الاحصاء الاستدلالي

مع تطبيقات

الطبعة الأولى 2022 مـ/1443هـ



دار المناهج للنشر والتوزيع

Dar Al-Manahej for Publishing & Distribution

عمان - الأردن -وسط البلد – شارع الملك حسين – بناية الشركة المتحدة للتأمين

هاتف: 00962 6 4650624 e-mail:manahej9@hotmail.com

المملكة الأردنية الهاشمية

رقم الإيداع لدى دائرة المكتبة الوطنية

(2021/8/4876)

518,54

البدران ، غضبان عبد الله خلف

الاحصاء الاستدلالي مع تطبيقات/غضبان عبد الله خلف البدران

_ عمان : دار المناهج للنشر والتوزيع ،2021

(ص).

ر.إ. : 2021/8/4876

الواصفات: الاستدلال الاحصائي/العينات/نظرية الاحتمالات/الاحصاء/

يتحمل المؤلف كامل المسؤولية القانونية عن محتوى مصنفه ولا يعبر هذا المصنف عن رأي دائرة المكتبة الوطنية أو أي جهة حكومية أخرى.

الرقم المعياري الدولي : 6 801 18 9957 978

الاحصاء الاستدلالي

مع تطبيقات

تأليف:

أ.د غضبان عبد الله البدران

أستاذ الاحصاء بكلية مزايا الجامعة

رئيس قسم هندسة تقنيات الحاسوب

كلية مزايا الجامعة /الناصرية -ذي قار



محتويات الفهرس

الفصل الثالث
اختبارات الفروض

الفصل الرابع: الاختبارات المعلمية

الفصل الخامس
مربع كاي (χ^2) واختبارات حسن المطابقة

الفصل السادس: تحليل التباين

الفصل السابع
تحليل السلاسل الزمنية

الفصل الثامن
الأرقام القياسية

المقدمة

يعتبر الإحصاء التحليلي واحدا من أهم التطبيقات لنظرية الإحصاء، حيث عن طريقه يمكن التعرف على الظواهر المختلفة والتنبؤ لما سيحدث لها في المستقبل وبالتالي فإن التقدم في جميع فروع العلم المختلفة يعتمد أساسا على هذا النوع من الإحصاء.

ويهدف عملنا هذا إلى التعرف على بعض طرق التحليل الإحصائي التي تعتمد على نظرية الإحصاء، حيث تم التعرض لبعض النظريات الإحصائية مبتعدا عن البراهين الرياضية راجيا أن أكون موفقا في إيصال المعلومة للقارئ بنوع من البساطة والسهولة بعيدا عن عمليات البراهين غير المطلوبة أساسا من أي باحث شريطة أن يكون لدى القارئ خلفية عن الإحصاء الوصفي. ولقد تعمدت أن أجعله عرضاً مبسطاً ليستفيد منه كل من تعنيه دراسة الإحصاء. وعلى الرغم من أني قد توسعت في بعض موضوعات الكتاب إلى الحد الذي يجاوز العرض المبسط، إلا أني وضعت نصب عيني حاجة الطالب في غالبية الموضوعات التي عالجتها.

وقد استعرض الكتاب بنوع من التحليل المفصل للاستدلال الإحصائي في العينات الكبيرة والصغيرة، كما تضمنت الدراسة في هذا الإطار كلا من الاختبارات الإحصائية وتقدير معالم المجتمع بالإضافة إلى نظرية النهاية المركزية.

ويمثل هذا الكتاب كتابا مكملا لكتابي الأول (طرق التحليل الإحصائي) والذي جاء متناولا للجانب الوصفي لعلم الإحصاء، وبذلك أرجو أن أكون قد قدمت للقارئ كتابا متكاملا في علم الإحصاء الوصفي والتحليلي، متمنيا أن يجد الباحث في كل مجال من المجالات الادارية أوالزراعية أو الصناعية أو الطبية أو علم النفس والاجتماع وغيرها من العلوم، أن يجد فيه الأداة العلمية للتحليل الإحصائي في كافة أوجه البحوث التجريبية والميدانية وحل العديد من المشاكل المتعلقة باتخاذ القرار.

أخيراً أقدم شكري وتقديري لمن ساهم معي في إخراج هذا الكتاب وبشكله الراهن. وأقدم شكري أيضاً لمن يلتمس لي العذر إن لاحظ قصوراً، فالشمول لا زال صعباً بشكل عام.

والله اسأل أن يأخذ بأيدينا لخدمة أهلنا دائماً

المؤلف

نيسان 2021

الفصل الأول

العينات والتوزيعات العينية

(1.1) العينات.

(2.1) توزيعات المعاينة.

(3.1) التوزيع العيني للمتوسط.

(4.1) نظرية النهاية المركزية.

(5.1) توزيع الفرق بين وسطين.

(6.1) التوزيع العيني للنسبة.

(7.1) التوزيع العيني للفرق بين نسبتين.

(8.1) التوزيع العيني للتباين.

(9.1) توزيع النسبة بين تبايني عينتين.

(10.1) ملخص للتوزيعات العينية.

العينات والتوزيعات العينية

Samples and Sampling Distributions

(1.1) العينات (*Samples*)

من المسائل المهمة التي تواجه الباحث العلمي عند شروعه في القيام بأي عمل هو تحديد نطاق العمل. وعلى الرغم من أن لكل بحث ظروفه الخاصة التي تحدد نطاق العمل اللازم له والمجهود والمال الممكن بذلهما في سبيل إنجازه، غير أنه من الواضح كلما زاد عدد المفردات التي يشملها البحث أيا كان نوعه كلما أصبحت النتائج التي يتوصل إليها الباحث تستند إلى أساس أمتن وأقوى. إلا أن مشقة العمل من جهة وإمكان الحصول على نتائج لا بأس بها بإجراء التجارب على عدد محدود من المفردات من الجهة الأخرى هو الذي جعل البحوث العلمية تقوم على أساس دراسة عينات محدودة مختارة من المجتمعات الإحصائية التي نرغب في التعرف على خواصها بدلا من دراسة المجتمعات الأصلية ذاتها (والذي يسمى بطريقة المسح الشامل *Census*)، ذلك لأنه في معظم الحالات يكاد يكون من المستحيل عمليا وصف المجموعات الكبيرة إلا عن طريق دراسة عينات تمثلها من الأخذ بنظر الاعتبار على أن محاولة التعرف على خواص المجموع عن طريق دراسة عينة مختارة منه تنطوي على بعض التضحية في دقة النتائج التي نحصل عليها. ويتوقف مقدار هذه التضحية إلى حد كبير على طريق اختيار العينة وحجمها التي سنجعلها موضع البحث.

وهناك أنواع عدة من العينات يستطيع الباحث أن يختار أحدها تبعا لظروف البحث الذي يجريه، **ومنها**:

- العينة العشوائية البسيطة *Simple Random Sample*
- العينة المقصودة *Purposive Sample*
- العينة الطبقية *Stratified Sample*
- العينة العنقودية *Cluster Sample*
- العينة المنتظمة *Systematic Sample*

وفيما يلي تعريف مع حالات استعمال الأنواع الثلاثة الأولى من العينات، وسوف تكون مناقشة العينة العشوائية البسيطة مفصلة نوعا ما وذلك للتعرف من خلالها على بعض المفاهيم الهامة مثل التوزيع العيني. أما الأنواع الباقية من العينات فهي متروكة إلى موضوع العينات الذي يناقش كل هذه الأنواع بالتفصيل.

أولا: العينة العشوائية البسيطة

على الرغم من أن اسم العينة العشوائية يحمل على الظن بأن مفرداتها تختار عشوائيا دونما قاعدة إلا أن الواقع خلاف ذلك. فالعينة العشوائية هي تلك التي تختار من المجموع بطريقة تعطي كل مفردة من مفرداته نفس الفرصة في الظهور. ويصلح هذا النوع من العينات للمجتمعات المتجانسة، بمعنى أنه إذا كان التوزيع الاحتمالي للمجتمع غير متجانس فإن العينة العشوائية البسيطة لا تصلح للاختيار.

وتعرف العينة العشوائية البسيطة على أنها طريقة لاختيار n من الوحدات من مجتمع حجمه N وحدة بحيث أن كل عينة من العينات لها نفس فرصة الاختيار حيث يطلق على العينة المختارة اسم "العينة العشوائية البسيطة" والسؤال هو ما هي الطريقة التي تتم بها اختيار هذه العينة؟ ببساطة إذا أردنا مثلا اختيار عينة عشوائية عدد مفرداتها 100 من بين طلاب كلية العلوم التي تضم 5000 طالب وحتى نضمن إعطاء الطلاب نفس فرصة الاختيار، نعطي كل طالب رقما معينا ونسجل هذه الأرقام على بطاقات متساوية الحجم تماما ثم نحكم خلطها جيدا ونختار منها المائة المطلوبة.

ومما تجب ملاحظته أنه على الرغم من أن الغرض من إعطاء كل مفردة نفس الفرصة في الظهور هو عدم التحيز نحو أي صفة من صفات المجموع دون الأخرى إلا أن هذا لا يعتبر في حد ذاته ضمانا كافيا لأن تكون العينة خالية تماما من التحيز وممثلة للمجموع تمثيلا دقيقا. والواقع أن دقة تمثيل العينة العشوائية للمجموع الذي اختيرت منه تتوقف على عدد مفردات العينة. حيث كلما كان هذا العدد كبيرا زادت دقة تمثيل العينة للمجموع.

ولهذا السبب يجب أن يقتصر استعمال العينة العشوائية على الحالات التي تسمح فيها ظروف الباحث باختيار عينات كبيرة حيث تكون عوامل المصادفة كفيلة بتحقيق الاختيار المحايد ما دامت لكل مفردة من مفردات المجموع نفس الفرصة بالظهور. ويطلق على هذا الأسلوب، الأسلوب العشوائي لاختيار العينة، وإن كان هناك طرقا متقدمة للاختيار العشوائي للعينة تدخل فيها الحاسبات الالكترونية أو جداول للاختيار العشوائي (جداول الأرقام العشوائية *Random Numbers*). وبموجب هذه الجداول يتم اختيار مثل هذه العينة بأن نعطي كل عنصر من عناصر المجتمع الإحصائي (N) رقما متسلسلا من صفر حتى $N\text{-}1$ بحيث تكون هذه الأرقام مكونة من نفس العدد من المنازل. فإذا رجعنا إلى مثال طلبة كلية العلوم في جامعة البصرة البالغ عددهم N=5000 طالبا، وأردنا اختيار عينة عشوائية بسيطة من بينهم فإننا نعطي الطلبة الأرقام المتسلسلة التالية والتي تتكون من أربعة مراتب: 0000, 0001, 0002 وهكذا حتى $N\text{-}1$ أي 4999. ثم نقوم باستعمال جدول الأرقام العشوائية الموجود في ملحق الكتاب ونقرأ من هذا الجدول عموديا بحيث يكون عدد مراتب كل عدد مساويا لعدد منازل الأرقام المتسلسلة في أعلاه. فإذا كان العدد الذي نقرأه من الجدول هو أحد الأرقام المتسلسلة قبلناه كعنصر من عناصر العينة وإلا رفضناه لننتقل إلى قراءة عدد آخر. ونستمر في القراءة كلما أنهينا عمودا انتقلنا إلى العمود الذي يليه حتى نحصل على عينة بالحجم المطلوب. فبالنسبة للمثال السابق إذا أردنا اختيار عينة حجمها 6

طلاب فإننا نحصل على الطلبة الذين أرقامهم: 1108, 4635, 4184, 2288, 3093, 3175.

ومن الجدير بالذكر أن هناك طريقتين للسحب يطلق على الأولى بطريقة السحب بالإرجاع والتي تعني أننا نسمح بإمكانية اختيار عنصر اختير في مرة سابقة، ويطلق على الأخرى بطريقة السحب بدون إرجاع والتي تعني أننا نرفض أي عدد أخذناه في قراءة سابقة.

ويجب ملاحظة الفروق بين احتمالات الأحداث في كل من الطريقتين ففي الطريقة الأولى تكون السحبات المتتالية مستقلة وأنها تخضع لتوزيع ذي الحدين (Binomial Distribution) في حين تخضع السحبات في الطريقة الثانية إلى توزيع فوق الهندسي (Hypergeometric Distribution) كونها سحبات متتالية غير مستقلة.

- **عدد العينات الممكنة**

لكي نتعرف على معلمة المجتمع المجهولة " وفي كل فروع الإحصاء المختلفة " فإننا نقوم بسحب عينة من المجتمع لدراسة خصائصها. هذه الخصائص يمكن أن تعطي معلومات كاملة عن المعلمة المجهولة تحت شروط معينة قابلة للتحقيق، مع ملاحظة أن الذي دفعنا إلى الاستعاضة عن دراسة المجتمعات الكبيرة بدراسة عينات مختارة هو مشقة العمل والوقت والتكاليف، غير أن مشقة العمل يجب ألا تتخذ عذرا لجعل العينات من الصغر. بحيث لا تتوافر فيها صفة تمثيل المجموع بدرجة معقولة من الدقة، فبدون توافر هذا الشرط يصبح البحث الذي نجريه لا قيمة له مطلقا بل ربما أصبح مضللا وخطيرا إذا بنيت آرائنا وتصرفاتنا عليه. فمن الواضح أن درجة الدقة تزداد كلما زاد عدد المفردات التي يشملها البحث أو بعبارة أخرى كلما زاد حجم العينة. فلا يجوز مثلا أن نعلن نظافة بلد يضم 30 مليون من السكان من وباء حل به على أساس فحص عشرات أو حتى مئات من الأفراد والتثبت من خلوهم من هذا الوباء، وإنما تكون نتائجنا أقرب إلى الدقة إذا كان بحثنا شاملا لأكثر من مليون فرد مثلا. وسوف نجد لاحقا أن هناك علاقة رياضية محددة بين حجم العينة ودقة النتائج التي نتوصل إليها

من دراسة هذه العينة. وبناءا على هذا كله علينا أن نتعرف على كيفية سحب العينة وعلى شروط سحبها لكي تمثل المجتمع أحسن تمثيل، كما يجب أن نتعرف على عدد العينات الممكن سحبها من المجتمع وعلى بعض الاحتمالات المتعلقة بعملية السحب.

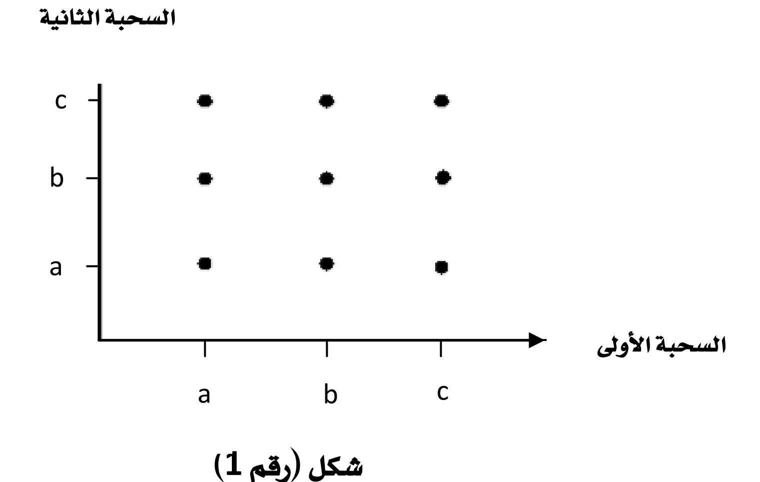
ويمكن توضيح ذلك كالآتي:

أ) عندما يكون السحب بالإرجاع

نفرض أن لدينا مجتمعا مكون من ثلاث وحدات $N=3$ حيث: $N_1 = a, N_2 = b, N_3 = c$ وأردنا سحب عينة من هذا المجتمع تتكون من مفردتين $n=2$. وحيث أن السحب بإرجاع فهناك إمكانية ظهور أي حرف من الحروف السابقة أكثر من مرة. وباستخدام قوانين التوافيق نجد أن عدد الطرق أو عدد جميع العينات الممكن اختيارها هو: $(3)^2 = 9$ عينة، موضحة كالتالي:

aa, ab, ac
ba, bb, bc
ca, cb, cc

والشكل التالي يمثل عدد العينات الممكنة من مجتمع مكون من ثلاث مفردات إذا كان حجم العينة يتكون من مفردتين وكان السحب بالإرجاع.



شكل (رقم 1)

وكقاعدة عامة

نفرض أن لدينا مجتمعا حجمه N سحبت منه عينه حجمها n، فإذا كان السحب بإرجاع **فان:**

1 - عدد العينات الممكن سحبها من هذا المجتمع هي: $(N)^n = 3^2 = 9$.

2 - احتمال سحب عينة حجمها n من هذا المجتمع هو: $\frac{1}{(N)^n} = \frac{1}{(3)^2} = \frac{1}{9}$.

ب) عندما يكون السحب بدون إرجاع

عندما يكون السحب بدون إرجاع فإنه سوف لا تكون هناك إمكانية لأن تظهر المفردة مرتين. فإذا كانت السحبة الأولى تحتوي على a فإن السحبة الثانية لن تحتوي على a، وهذا يعني أننا لن نجد العينات aa, bb, cc. كذلك إذا تم سحب العينة ab فسوف لن نجد العينة ba وهكذا. وبذلك تكون العينات الممكن سحبها من ذلك المجتمع والتي تتكون من مفردتين هي: ab, ac, bc.

وكقاعدة عامة

نفرض أن حجم المجتمع يتكون من N مفردة وسحبنا عينة منها بحجم n حيث كان السحب بدون إرجاع فإن:

1 -عدد جميع العينات الممكن سحبها من ذلك المجتمع هو عدد الطرق التي يمكن بها أن نختار

n مفردة من عدد من المفردات قدرها N، وهذا العدد هو:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{3!}{2!(3-2)!} = 3$$

2 - احتمال سحب عينة حجمها n من ذلك المجتمع هو :

$$\frac{1}{\binom{N}{n}} = \frac{1}{\binom{3}{2}} = \frac{1}{3}$$

ثانياً: العينة المقصودة

إذا كانت ظروف البحث لا تسمح إلا باختيار عدد قليل من المفردات فان صفة تمثيل المجموع تمثيلا دقيقا لا تتوافر في العينة العشوائية كما أسلفنا والأفضل أن نختار هذه المفردات القلائل بطريقة يراعى فيها أن تكون جميع المفردات قريبة من المتوسط كي تتوافر فيها صفة التمثيل المطلوبة، وتسمى العينة المختارة بهذه الكيفية عينة مقصودة.

ومن أهم عيوبها أنها لا تختار بطريقة موضوعية تستبعد اثر العنصر الشخصي من الاختيار كما هو الحال في العينة العشوائية ولذلك فمن السهل أن تأتي متحيزة نحو صفة ما من صفات المجموع. وهي لهذا السبب ولغيره يجب ألا نلجأ إلى استعمالها إلا إذا اضطرتنا ظروف البحوث إلى اختيار عدد قليل جدا من المفردات.

ثالثاً: العينة الطبقية

في كثير من الحالات يكون المجتمع الذي نختار منه إما منقسما بطبيعته إلى طبقات أو يمكن تقسيمه بسهوله إلى طبقات يتصف كل منها بصفة خاصة تميزها عن غيرها. فإذا كنا نحرص على إبراز هذه الصفات جميعا في العينة، فإننا نعتبر كل طبقة من طبقات المجتمع مجتمعا مستقلا ونختار من كل منها عينة عشوائية. فإذا جمعنا هذه العينات معا لنكون من بينها عينة واحدة فإننا نحصل على ما يسمى بالعينة الطبقية. ويلاحظ أن هذه الطريقة تعتبر وسطا بين الطريقتين السابقتين فهي تجمع بعضا من مزايا كل منها، ففيها عنصر الاختيار المتعمد الذي يضمن عدم خلو العينة من بعض الصفات الهامة الموجودة في المجتمع وفيها كذلك عنصر الاختيار المحايد الذي يقلل احتمال التحيز نحو مفردة أو أخرى داخل كل طبقة.

ولتوضيح طريقة اختيار العينة الطبقية، افرض مثلا إننا أردنا سحب عينة طبقية من ذوي الحيازة الزراعية في محافظة معينة. وللحصول على هذه العينة نبدأ بتقسيم الملاك إلى طبقات حسب المساحة التي يملكونها كما يلي:

(أقل من فدان، 5 - 1 ، 10 - 5 ،...) ثم نختار من كل طبقة عينة عشوائية بسيطة. وبذلك نظمن أن تكون العينة الطبقية المكونة من هذه العينات شاملة لكافة طبقات الملاك الزراعيين.

ولاشك من أن العينة الطبقية أفضل من العينات الأخرى في تلك الحالات التي يكون المجتمع فيها منقسما بطبيعته إلى أقسام مميزة عن بعضها ويمثل كل قسم منها وحدة متجانسة من المفردات تقريبا. أما إذا كان المجتمع كله متجانسا أو قريبا من التجانس فلا مجال لاستخدام العينة الطبقية.

(2.1) توزيعات المعاينة (*Sampling Distributions*)

بعد أن ينتهي الباحث من تحديد أهداف الدراسة الإحصائية وينتهي من سحب عينة عشوائية من المجتمع الإحصائي المحدد، فإنه يقوم بحساب إحصائيات هذه العينة مثل المتوسط والانحراف المعياري والنسبة، كتقديرات لمعالم المجتمع الأصلي أي μ, σ, p . والسؤال هو ما هي طبيعة العلاقة بين التوزيع الأصلي للمجتمع بمعالمه المختلفة وتوزيع المتوسطات الحسابية للعينات التي يمكن اختيارها عشوائيا من هذا المجتمع بإحصائياته المختلفة. إن توزيعات الإحصاءات المختلفة تسمى بالتوزيعات العينية (المعاينة)، وما سنقوم بتفصيله الآن هو التعرف على هذه التوزيعات المختلفة.

(3.1) التوزيع العيني للمتوسط

(*Sampling Distribution of the Sample Mean*)

في الحياة العامة غالبا ما يكون حجم المجتمع كبيرا وكذلك عدد العينات مما يصعب معه كمية المعلومات التي يمكن أن تستنتج من العينات المختلفة وكيفية استخدامها في التعرف على خصائص المجتمع. وسوف نتعرف على كل

هذا من خلال مثال لمجتمع عدد مفرداته صغير لنصل إلى أحكام عامة تنطبق على أي حجم للمجتمع وعلى أي حجم للعينة.

مثال (1): نفرض أن لدينا مجتمعا إحصائيا مكونا من ثلاثة وحدات (*N=3*)، وهي: $X_1 = 1, X_2 = 2$

$X_3 = 3$ ، وقبل البدء في سحب العينات يمكن أن نتعرف على بعض معالم ذلك المجتمع كمتوسطه وتباينه كما يلي:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{6}{3} = 2$$

$$\sigma^2 = \frac{1}{N}(\sum_i X_i^2 - \frac{(\sum_i X_i)^2}{N}) = \frac{2}{3}.$$

(أ) التوزيع العيني في حالة السحب بالإرجاع

أفرض أننا سحبنا عينة عشوائية من ذلك المجتمع مكونة من مفردتين حيث كان السحب بالإرجاع فإن عدد العينات الممكنة سوف يكون مكونا من 9 عينات، والجدول التالي يمثل جميع العينات من الحجم 2 (أي 2= *n*).

جدول (رقم 1)

رقم العينة	1	2	3	4	5	6	7	8	9
مفرداتها	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3

من الجدول أعلاه يتضح أنه يمكن حساب المتوسط لكل عينة من العينات المسحوبة، وسوف تكون قيمة ذلك المتوسط متغيرة وعشوائية لأنها تختلف من محاولة إلى أخرى وأن قيمته تتوقف على الصدفة وبذلك فإن المتوسط يصبح متغيرا عشوائيا يمكن دراسته دراسة كافية بالحصول على دالة التوزيع الاحتمالي الخاصة به وكذلك على مقاييس النزعة المركزية والتشتت له.

والجدول التالي يوضح حساب متوسط كل عينة من العينات التسع السابقة مع بيان احتمال سحب كل عينة من هذه العينات.

جدول (رقم 2)

العينة	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
$\bar{X}_i$	1	1.5	2	1.5	2	2.5	2	2.5	3
$p(\bar{X})$	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

ومن أعلاه يمكن إيجاد التوزيع للمتغير العشوائي $\bar{X}_i$ والذي يطلق عليه أسم التوزيع الاحتمالي (أو العيني) للمتوسط في حالة السحب بالإرجاع.

والجدول الآتي يبين هذا التوزيع مع بعض الحسابات الخاصة بإيجاد توقع وتباين $\bar{X}$ كمقياسين للنزعة المركزية والتشتت على التوالي.

جدول (رقم 3)

التوزيع العيني للمتوسط مع حساب توقع وتباين $\bar{X}$

في حالة السحب بالإرجاع

$\bar{X}_i$	$p(\bar{X}_i)$	$\bar{X}\ p(\bar{X})$	$(\bar{X}-E(\bar{X}))^2\,p(\bar{X})$
1	1/9	1/9	1/9
1.5	2/9	6/18	2/36
2	3/9	6/9	0
2.5	2/9	10/18	2/36
3	1/9	3/9	1/9
المجموع	**1**	**2**	**1/3**

وليس من الصعب إثبات أن التوزيع العيني في الجدول أعلاه يمثل توزيعا احتماليا حيث أن مجموع الاحتمالات تساوي الواحد الصحيح (أي $\sum p(\overline{X})=1$) كما أن $p(\overline{X})\geq 0$.

ومن الجدول كذلك تسجل الملاحظات الآتية:

أ) متوسط التوزيع العيني هو توقع $\overline{X}$ أي أن:

$$E(\overline{X})=\sum \overline{X}\ p(\overline{X})=2$$

وتؤكد هذه النتيجة على أن هناك علاقة بين متوسط التوزيع العيني $(E(\overline{X}))$ في حالة السحب بالإرجاع وبين متوسط المجتمع μ بعد أن تم حسابه ووجد أنه يساوي 2 أيضا، وهذا يعني أن:

$$\mu = E(\overline{X})$$

ب) تباين التوزيع العيني هو:

$$Var(\overline{X})=\sum (\overline{X}-E(\overline{X}))^2\, p(\overline{X})=1/3$$

وقد وجد أن هناك علاقة بين هذا المقدار وبين تباين المجتمع حيث وجد أن:

$$Var(\overline{X})=\sigma^2/n$$

ويمكن التأكد من ذلك حسابيا كما هو أدناه بعد أن نتذكر أن تباين المجتمع σ^2 يساوي 2/3 :

$$Var(\overline{X})=\frac{\sigma^2}{n}=(\frac{2}{3})(\frac{1}{2})=\frac{1}{3}$$

وهي نفس النتيجة السابقة.

(ب) التوزيع العيني في حالة السحب بدون إرجاع

بنفس الأسلوب السابق، يمكن الحصول أيضا على التوزيع العيني في حالة السحب بدون إرجاع، حيث يوجد لدينا والحالة هذه ثلاث عينات هي: (2,3) ,(1,3) ,(1,2) ثم يحسب المتوسط لكل عينة وحيث أن احتمال سحب كل عينة هو 1/3، فانه يمكن تمثيل التوزيع العيني للمتوسط مع بعض الحسابات الضرورية لإيجاد توقع وتباين $\bar{X}$ في الجدول التالي.

جدول (رقم 4)

التوزيع العيني للمتوسط مع حساب توقع وتباين $\bar{X}$

في حالة السحب بدون إرجاع

العينة	$\bar{X}_i$	$p(\bar{X}_i)$	$\bar{X}\ p(\bar{X})$	$(\bar{X}-E(\bar{X}))^2\ p(\bar{X})$
1, 2	1.5	1/3	3/6	1/12
1, 3	2.0	1/3	2/3	0
2, 3	2.5	1/3	5/6	1/12
المجموع		**1**	**2**	**1/6**

ومن الجدول أعلاه نلاحظ ما يأتي:

(أ) $E(\bar{X})=\mu$

(ب) $$Var(\bar{X})=\frac{\sigma^2}{n}.\frac{N-n}{N-1}=(\frac{2}{3}.\frac{1}{2})(\frac{3-2}{3-1})=\frac{1}{6}$$

وهي نفس النتيجة المتحصل عليها من التوزيع العيني. ويسمى المقدار $\frac{N-n}{N-1}$ بمعامل التصحيح للمجتمعات المحدودة وهو يساوي الواحد الصحيح عندما تؤول N إلى ∞، فعندما يكون حجم المجتمع كبيرا جدا حينها لا يوجد فرق بين السحب بإرجاع والسحب بدون إرجاع.

ومن هذا المثال يمكن أن نستنتج الآتي:

(1) يمكن إيجاد التوزيع العيني للمتغير $\bar{X}$ بأن نحصل على كل العينات المكنة من المجتمع سواء كان السحب بإرجاع أو بدون إرجاع ثم نحسب المتوسط لكل عينة، وحيث أن احتمال سحب العينة في حالة الإرجاع وعدم الإرجاع معروف فانه يمكن إيجاد دالة التوزيع الاحتمالي للمتغير $\bar{X}$.

(2) متوسط التوزيع العيني للمتوسط وتباينه في حالة السحب بالإرجاع أو بدون إرجاع يكونان كالتالي:

$$(a) \qquad E(\bar{X}) = \mu$$

$$(b) \qquad Var(\bar{X}) = \begin{cases} \dfrac{\sigma^2}{n}, & with\ replacement \\ \dfrac{\sigma^2}{n} . \dfrac{N-n}{N-1}, & without\ replacement \end{cases}$$

لاحظ أن تشتت التوزيع العيني في حالة السحب بدون إرجاع أقل منه في حالة السحب بالإرجاع ويتساوى تشتت التوزيعين عندما تكون n كبيرة جدا، حيث كلما زاد حجم العينات كلما قل الانحراف المعياري لمتوسطاتها وكلما أصبحت النتائج التي نصل إليها بإحلال الوسط الحسابي للعينة محل الوسط الحسابي للمجتمع الأصلي أقرب إلى الدقة.

- **مثال (2):** مجتمع إحصائي يتكون من الأعداد 4, 5, 3, 8 ، سحبت منه عينة حجمها 3 بالإرجاع مرة وبدون إرجاع مرة أخرى. أوجد: $E(\bar{X}),\ \ Var(\bar{X})$.

الحل:

$$E(\bar{X}) = \mu = \frac{4+5+3+8}{4} = 5$$

وهو كذلك، سواء كان السحب بإرجاع أو بدون إرجاع. وقبل حساب تباين $\bar{X}$ في كلا الحالتين يجب أن نحسب تباين المجتمع σ^2 كما يلي:

$$\sigma^2 = \frac{1}{N}\sum_i (X_i - \mu)^2 = \frac{1}{4}\left[(4-5)^2 + (5-5)^2 + (3-5)^2 + (8-5)^2\right] = 3.5$$

وفي حالة السحب بالإرجاع فان تباين $\bar{X}$ هو:

$$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{3.5}{3} = \frac{7}{6}$$

بينما يكون التباين في حالة السحب بدون إرجاع كما يلي:

$$Var(\bar{X}) = \frac{\sigma^2}{n}.\frac{N-n}{N-1} = (\frac{3.5}{3})(\frac{4-3}{4-1}) = \frac{7}{18}$$

واضح أن تباين $\bar{X}$ في حالة السحب بدون إرجاع أقل منه في حالة السحب بالإرجاع.

(4.1) نظرية النهاية المركزية (*The Central Limit Theorem*)

من الثابت انه إذا كان المجتمع الأصلي المسحوبة منه العينات يتبع التوزيع االطبيعي (Normal Distribution) فان التوزيع العيني للمتوسط سوف يكون معتدلا هو الآخر بمتوسط قدره μ في حالة الإرجاع أو عدم الإرجاع، وتباين $\frac{\sigma^2}{n}$ عندما يكون السحب بالإرجاع، و $Var(\bar{X}) = \frac{\sigma^2}{n}.\frac{N-n}{N-1}$ عندما يكون السحب بدون إرجاع. ويمكن حينها تطبيق جميع الخصائص الخاصة بالتوزيع المعتدل على التوزيع العيني للمتوسط. وهذا ما تلخصه النظرية الآتية:

نظرية (1)

إذا كانت $x_1, x_2, \ldots, x_n$ عينة عشوائية وسطها $\overline{X}$ مسحوبة من توزيع طبيعي وسطه μ وتباينه σ^2 فان المتغير العشوائي Z :

$$Z = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}}$$

يتبع التوزيع المعتدل المعياري بوسط يساوي صفر وتباين يساوي واحد.

إما إذا كان المجتمع الأصلي لا يتبع التوزيع المعتدل فقد تمت برهنة نظرية غاية في الأهمية من الوجهة الإحصائية تحدد شكل التوزيع للمتوسط إذا كان التوزيع الاحتمالي للمجتمع معتدلا أو غير معتدل. هذه النظرية يطلق عليها اسم نظرية الحد (النهاية) المركزية، والتي يمكن تلخيصها كالآتي:

إذا كان X متغيرا عشوائيا توزيعه $p(x)$ ، وهذا التوزيع قد يكون معتدلا أو لا يكون وقد يكون متماثلا أو لا يكون وربما كان متقطعا أو مستمرا. نفرض أن متوسط X هو μ وتباينه $\sigma^2, (\sigma^2 \neq \infty)$. والآن إذا سحبنا كل العينات من الحجم n فإن التوزيع العيني لمتوسط العينات المسحوبة من ذلك المجتمع يؤول إلى التوزيع المعتدل بمتوسط μ وتباين $\frac{\sigma^2}{n}$ شرط أن يكون حجم العينة كبيرا كبرا كافيا ومن هنا يتضح أن المتغير Z:

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

يؤول إلى التوزيع المعتدل المعياري بغض النظر عن شكل التوزيع الخاص بالمتغير X بشرط أن تكون n كبيرة. أما الحد الذي يمكن معه اعتبار أن n كبيرة جدا أو كبيرة كبرا كافيا يسمح بتطبيق النظرية فهو عندما تكون $n \geq 30$. والنظرية التالية تلخص ما سبق.

نظرية (2)

إذا سحبت عينة حجمها n من مجتمع إحصائي وسطه μ وتباينه σ^2 ، فإن $\bar{X}$ يخضع تقريبا لتوزيع طبيعي وسطه μ وتباينه $\frac{\sigma^2}{n}$ عندما تكون n كبيرة كبرا كافيا، أي أن $Z=\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma}$ تخضع لتوزيع طبيعي معياري $N(0,1)$

مع الأخذ بنظر الاعتبار الملاحظات التالية:

أ) إذا سحبت العينة من توزيع طبيعي أصلا وسطه μ وتباينه σ^2 فإن نتيجة النظرية أعلاه سوف تكون دقيقة وليست مقربة.

ب) إذا كانت σ^2 مجهولة فإننا نستخدم الانحراف المعياري للعينة S بدلا من الانحراف المعياري للمجتمع σ وعندها يكون المتغير $Z=\frac{(\bar{X}-\mu)\sqrt{n}}{S}$ قريبا من التوزيع الطبيعي إذا كان حجم العينة كبيرا أي $(n \geq 30)$. أما إذا كان حجم العينة صغيرا فإن المتغير Z أعلاه لا يكون قريبا من التوزيع الطبيعي.

ج) إذا كان حجم العينة صغيرا $(n < 30)$ وكان المجتمع خاضعا لتوزيع طبيعي فإن المتغير:

$$t=\frac{(\bar{X}-\mu)\sqrt{n}}{S}$$

يخضع لتوزيع t بدرجات حرية $n-1$. ويمكن توضيح النظرية بالمثال التالي.

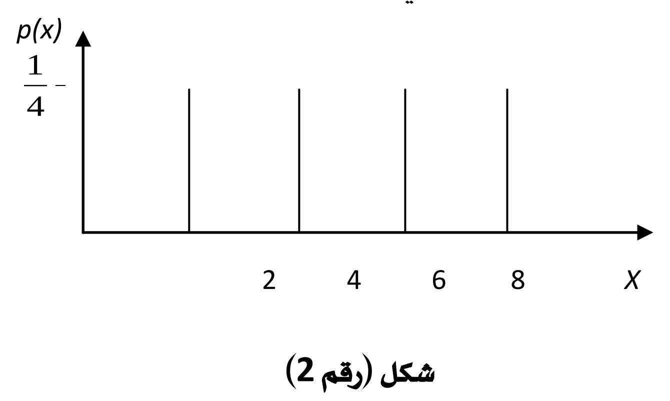
- **مثال (3):** أفرض أن لدينا مجتمعا مكونا من 4 مفردات حيث $X_1=2, X_2=4, X_3=6, X_4=8.$ تحقق من أن التوزيع العيني للمتوسط يقترب من التوزيع المعتدل كلما كبر حجم العينة.

الحل: يمكن حساب معالم المجتمع من هذه البيانات كما يلي:

$$\mu = \frac{\sum X}{N} = \frac{20}{4} = 5.$$

$$\sigma^2 = \frac{1}{N}(\sum X^2 - N\mu^2) = \frac{1}{4}(120 - 4(5)^2) = 5.$$

واضح جدا أن توزيع المتغير X لا يتبع التوزيع المعتدل بل يتبع توزيعا آخر يطلق عليه التوزيع المنتظم (*Uniform Distribution*) وهو شكل المجتمع المسحوبة منه العينة والذي يمكن تمثيله بالشكل التالي:



شكل (رقم 2)

لنفرض الآن أننا نريد سحب عينة من الحجم 2 بالإرجاع ثم نوجد التوزيع العيني ونرسمه لنحدد هل سيقترب التوزيع من المنحنى المعتدل أم لا؟ في حالة السحب بالإرجاع فإن عدد العينات الممكنة هي: $N^n = 4^2 = 16$ عينة، وبيانها كالتالي:

2,2	2,4	2,6	2,8
4,2	4,4	4,6	4,8
6,2	6,4	6,6	6,8
8,2	8,4	8,6	8,8

كما نعلم، فإن احتمال سحب أي عينة من هذه العينات هي 1/16. وبحساب متوسط كل عينة فإننا نحصل على جدول التوزيع العيني للمتوسط.

جدول (رقم 5)

التوزيع العيني للمتوسط $\overline{X}$ $(n = 2)$

$\overline{X}$	2	3	4	5	6	7	8
$p(\overline{X})$	1/16	2/16	3/16	4/16	3/16	2/16	1/16

وإذا قمنا بتمثيل الجدول أعلاه بيانيا فإننا نلاحظ أن التوزيع العيني للمتوسط يقترب من التوزيع المعتدل، كما إننا نجد:

$$E(\overline{X}) = \sum(\overline{X})\ p(\overline{X}) = \mu = 5$$

$$Var(\overline{X}) = \frac{\sigma^2}{n} = \frac{5}{2} = 2.5$$

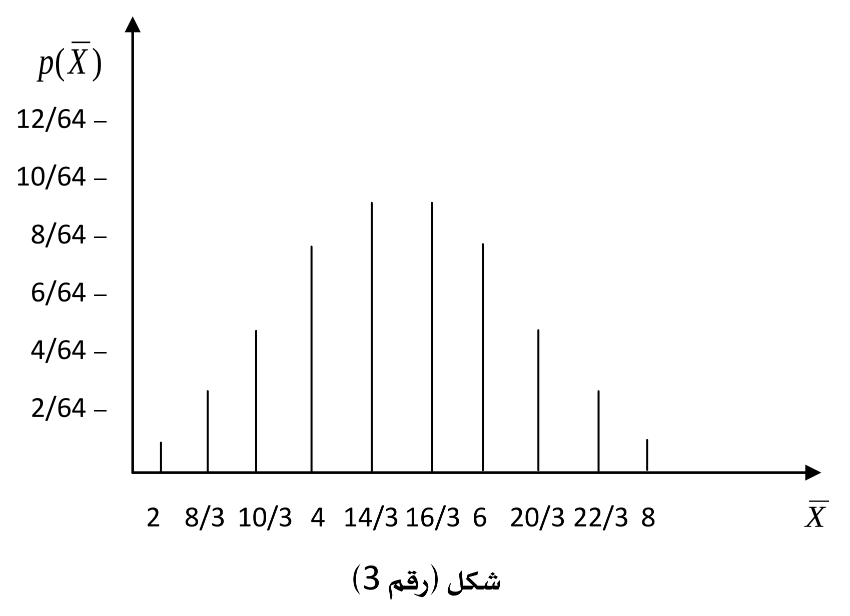
أي أنه عندما يكون المجتمع الأصلي والذي سحبت منه العينات غير معتدل فإن التوزيع العيني للمتوسط يقترب من التوزيع المعتدل بمتوسط $\mu = 5$ وتباين يساوي $\frac{\sigma^2}{n} = 2.5$ وسوف ينطبق على المنحنى المعتدل عندما تكون $n > 30$.

ولكي نراقب اقتراب التوزيع العيني للمتوسط من التوزيع المعتدل نفرض أننا سحبنا كل العينات من الحجم $n = 3$. وفي هذه الحالة يكون عدد العينات الممكن سحبها بإرجاع هي: $4^3 = 64$ عينة. وبنفس الأسلوب السابق يمكن أن نحصل على الجدول أدناه وهو خاص بالتوزيع العيني للمتوسط عندما تكون $n = 3$.

جدول (رقم 6)

$\overline{X}$	2	8/3	10/3	4	14/3	16/3	6	20/3	22/3	8
$p(\overline{X})$	1/64	3/64	6/64	10/64	12/64	12/64	10/64	6/64	3/64	1/64

وبتمثيل الجدول أعلاه فإننا نحصل على الشكل (رقم 3) وهو يقترب من التوزيع المعتدل أكثر من اقتراب الجدول (رقم 4) في حالة تمثيله.



شكل (رقم 3)

وإذا حسبنا $E(\bar{X})$, $Var(\bar{X})$ من الجدول السابق نجد أن:

$$E(\bar{X}) = \mu = 5, \qquad Var(\bar{X}) = \frac{\sigma^2}{n} = 2.5$$

وبذلك نكون قد تحققنا من النظرية.

- **مثال (4)**: إذا كان X متغيرا عشوائيا يمثل أوزان 49 طالبا قيست أوزانهم فكان وسطها 65 كغم وانحرافها المعياري 10.5 كغم. احسب احتمال أن يكون المتوسط أكبر من 67.5 كغم؟

الحل:

$$p(\bar{X} > 67.5) = p(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{67.5 - \mu}{\frac{\sigma}{\sqrt{n}}})$$

$$= p(\frac{\bar{X} - 65}{10.5/7} > \frac{67.5 - 65}{10.5/7}) = p(Z > 1.67)$$

وباستخدام جدول التوزيع الطبيعي المعياري (جدول رقم 2) نجد أن:

$$p(\bar{X} > 67.5) = p(Z > 0) - p(0 < Z < 1.67) = 0.5 - 0.4525 = 0.0475$$

- **مثال (5):** إذا كانت أطوال طلاب احدى الكليات تخضع لتوزيع طبيعي وسطه 168 سم. سحبت عينة حجمها 25 طالبا من هذه الكلية فكان الانحراف المعياري لأطوال طلاب هذه العينة هو 12 سم. احسب احتمال أن يكون الوسط الحسابي لأطوال طلاب هذه العينة اكبر من 174 سم.

الحل:

لأن حجم العينة صغير وأن تباين المجتمع σ مجهول وأن توزيع X طبيعي فإن المتغير:

$$t = \frac{(\bar{X} - \mu)\sqrt{n}}{S} = \frac{\bar{X} - 168}{12/5} \quad ,$$

تخضع لتوزيع t بدرجات حرية $n - 1 = 24$

$$p(\bar{X} > 174) = p(\frac{\bar{X} - 168}{12/5} > \frac{174 - 168}{12/5})$$
$$= p(t > 2.5)$$

وباستعمال جدول t في الملحق نجد أن الاحتمال المطلوب هو:

$$p(\bar{X} > 174) = 0.01 .$$

- **مثال (6):** نفرض أن لدينا مجتمعا مكونا من 6 مفردات ، حيث $X_1 = 1, X_2 = 2,$ $X_3 = 3, X_4 = 4, X_5 = 5, X_6 = 6$. ويفرض أننا سحبنا عينة من وحدتين بدون إرجاع فإن عدد العينات التي يمكن سحبها من ذلك المجتمع هو: $\binom{6}{2} = 15$ ، والتي يمكن بيانها كما يلي:

1,2   1,3   1,4   1,5   1,6
2,3   2,4   2,5   2,6
3,4   3,5   3,6
4,5   4,6
5,6

من المعلوم أن احتمال سحب أي عينة من هذه العينات هو 1/15. وبحساب متوسط كل عينة فإنه يمكن الحصول على التوزيع العيني للمتوسط كما في الجدول التالي:

جدول (رقم 7)

$\bar{X}$	1.5	2	2.5	3	3.5	4	4.5	5	5.5
$p(\bar{X})$	1/15	1/15	2/15	2/15	3/15	2/15	2/15	1/15	1/15

ومن الجدول أعلاه يمكن حساب ما يأتي:

$$E(\bar{X}) = \sum \bar{X} \; p(\bar{X})$$

$$= 1.5(1/15) + \ldots + 5.5(\frac{1}{15})$$

$$= 3.5$$

$$Var(\bar{X}) = \sum (\bar{X} - E(\bar{X}))^2 \, p(\bar{X})$$

$$= (1.5 - 3.5)^2 (\frac{1}{15}) + \ldots + (5.5 - 3.5)^2 (\frac{1}{15})$$

$$= 17.5/15$$

وبمقارنة هذه البيانات مع بيانات المجتمع التي يمكن حسابها كما يلي:

$$\mu = \frac{\sum X}{N} = \frac{21}{6} = 3.5$$

$$\sigma^2 = \frac{1}{N} \sum (X - \mu)^2 = \frac{17.5}{6}$$

فإن المقارنة تظهر أن التوزيع العيني للمتوسط في حالة السحب بدون إرجاع يتوزع توزيعا قريبا من التوزيع المعتدل بمتوسط يساوي μ وتباين $\frac{\sigma^2}{n}.\frac{N-n}{N-1}$، حيث أن:

$$Var(\overline{X}) = (\frac{17.5}{6})(\frac{1}{2})(\frac{6-2}{6-1})$$
$$= \frac{17.5}{15}$$

وهو نفس التباين المحسوب من التوزيع العيني للمتوسط في جدول (رقم 6).

- **مثال (7)**: في مثال (6) احسب احتمال أن يكون المتوسط أكبر من أو يساوي 4.5 وذلك باستخدام التوزيع العيني للمتوسط والتوزيع المعتدل وقارن بينهما؟

الحل:

في هذا المثال نحاول إيجاد المساحة تحت المنحنى المعتدل الذي يستخدم كتقريب للتوزيع العيني وهذه المساحة تمثل احتمال تقريبي. أما حساب الاحتمال من التوزيع العيني فهو يمثل احتمالا مضبوطا، وبالتالي:

1 -باستخدام دالة التوزيع الاحتمالي للمتوسط (جدول رقم 7) فإن الاحتمال المطلوب هو:

$$p(\overline{X} \geq 4.5) = \frac{2}{15} + \frac{1}{15} + \frac{1}{15} = \frac{4}{15} = 0.267$$

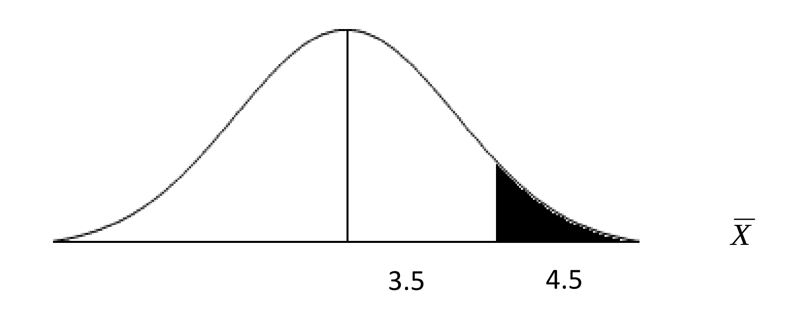
وهذا هو الاحتمال الحقيقي والمضبوط.

2 -باستخدام نظرية النهاية المركزية، فإن $\overline{X}$ يتوزع في هذا المثال توزيعا معتدلا عياريا

بمتوسط 3.5 وانحراف معياري: $\sqrt{\frac{17.5}{15}} = 1.08$، وهذا يعني أن:

$$Z = \frac{\overline{X} - 3.5}{1.08} \sim N(0,1)$$

وباستخدام هذا التقريب فإن احتمال أن يكون $\overline{X}$ أكبر من أو يساوي 4.5 يمكن مشاهدته على الشكل التالي:



شكل (رقم 4)

وهو يمثل المساحة المضللة. وباستخدام خصائص المنحنى المعتدل يمكن إيجاد المساحة المطلوبة كما يلي:

$$p(\overline{X} > 4.5) = p(\frac{\overline{X} - \mu}{1.08} > \frac{4.5 - \mu}{1.08})$$
$$= p(\frac{\overline{X} - 3.5}{1.08} > \frac{4.5 - 3.5}{1.08})$$
$$= p(Z > 0.93)$$
$$= 0.5 - p(0 < Z < 0.93)$$
$$= 0.5 - 0.3238$$
$$= 0.1762$$

واضح جدا أن هناك اختلافا بين الاحتمالين السابق حسابهما من التوزيع العيني ومن التوزيع المعتدل وهذا راجع إلى أن n ليست كبيرة بما فيه الكفاية.

(5.1) توزيع الفرق بين وسطين
(*Distribution of the Difference Between Two Sample Means*)

نحتاج وفي أحيان كثيرة إلى توزيع الفرق بين وسطين لعينتين. وهذا التوزيع تعطيه النظرية التالية إذا كانت العينتان خاضعتين لتوزيعين طبيعيين مستقلين والتباينات معلومة.

نظرية (3)

أ) إذا كانت $X_1,\ldots,X_n$ عينة وسطها $\bar{X}$ مأخوذة من توزيع طبيعي $N(\mu_1,\sigma_1^2)$، وكانت $Y_1,\ldots,Y_m$ عينة أخرى وسطها $\bar{Y}$ مستقلة عن العينة الأولى مأخوذة من توزيع طبيعي آخر $N(\mu_2,\sigma_2^2)$، فإن المتغير العشوائي Z يخضع لتوزيع طبيعي معياري $N(0,1)$ بغض النظر عن أحجام العينات، أي أن:

$$Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n}+\dfrac{\sigma_2^2}{m}}}\sim N(0,1)$$

ب) وعندما تكون التباينات غير معلومة ولكنها متساوية (أي أن $\sigma_1^2=\sigma_2^2=\sigma^2$) وكانت حجوم العينات صغيرة ($n+m<50$)، فقد وجد أن المتغير:

$$t=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{S_p^2(\dfrac{1}{n}+\dfrac{1}{m})}}$$

يخضع لتوزيع t بدرجات حرية ($n+m-2$) حيث S_p^2 هو التباين الموحد (*Pooled Variance*) للعينتين والذي تعرفه العلاقة الآتية:

$$S_p^2=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}$$

وفي حالة كون التباينات غير متساوية ($\sigma_1^2 \neq \sigma_2^2$) فقد وجد أن المتغير:

$$t' = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_{2)}}{\sqrt{S_1^2/n+S_2^2/m}}$$

يخضع تقريبا لتوزيع t بدرجات حرية تساوي

$$d.f = \frac{(\frac{S_1^2}{n}+\frac{S_2^2}{m})^2}{\frac{(S_1^2/n)^2}{n-1}+\frac{(S_2^2/m)^2}{m-1}}$$

أما إذا كانت العينتان غير مسحوبتين من مجتمعين خاضعين للتوزيع الطبيعي، فإن النظرية التالية تحدد التوزيع للفرق بين الوسطين وذلك بالاستفادة من نظرية النهاية المركزية.

نظرية (4)

(أ) إذا سحبت عينة كبيرة حجمها n ووسطها $\bar{X}$ من مجتمع وسطه μ_1 وتباينه σ_1^2 معلوم، ثم سحبت عينة أخرى كبيرة حجمها m ووسطها $\bar{Y}$ من مجتمع آخر وسطه μ_2 وتباينه σ_2^2 معلوم مستقل عن المجتمع الأول فإن المتغير Z يخضع تقريبا لتوزيع طبيعي معياري أي أن:

$$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1)$$

(ب) أما إذا كانت التباينات غير معلومة وكان حجم العينات كبيرا ($n + m > 50$)، فإنه يمكن استعمال S_1^2, S_2^2 (التباين المحسوب من كل عينة) كتقديرين لكل من σ_1^2, σ_2^2 على التوالي، وبذلك يكون المتغير Z:

$$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{S_1^2/n+S_2^2/m}}$$

يخضع تقريبا لتوزيع طبيعي معياري $N(0,1)$.

- **مثال (8)**: كانت درجات طلبة كلية ما في جامعة بغداد تخضع لتوزيع طبيعي $N(70, 0.5)$. وفي جامعة الملك البصرة كانت درجات الطلبة لنفس الكلية تخضع لتوزيع طبيعي $N(69,0.25)$ ، فإذا أخذت عينة من الجامعة الأولى حجمها 16 طالبا وأخرى من الجامعة الثانية حجمها 8 طلاب، فالمطلوب حساب أن احتمال الفرق بين متوسطي العينتين أقل من أو يساوي 0.80؟

الحل: إذا كان $\bar{X}$ يرمز لوسط درجات العينة الأولى، وكان $\bar{Y}$ يرمز لوسط درجات العينة الثانية فإن المطلوب هو $p(\bar{X}-\bar{Y}\le 0.80)$ وبالتالي:

$$p(\bar{X}-\bar{Y}\le 0.80)=p\left(\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\le\frac{0.80-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\right)$$

$$=p\left(Z\le\frac{0.80-(70-69)}{\sqrt{\frac{0.5}{16}+\frac{0.25}{8}}}\right)$$

$$=p(Z\le -0.2)$$

$$=0.5-p(0\le Z\le 0.2)$$

$$=0.4207$$

(6.1) التوزيع العيني للنسبة

The Sampling Distribution of the sample Proportion

للتعرف على المقصود بكلمة النسبة نفرض أن لدينا مجتمعا مكونا من N وحدة ثم قمنا بمعرفة عدد الذين يملكون خاصية في ذلك المجتمع كعدد المدخنين مثلا. من هذه البيانات يمكن إيجاد نسبة المدخنين في المجتمع وتعتبر هذه النسبة أحد معالم المجتمع المجهولة وهدفنا هو التعرف عليها وتقديرها من العينة.

وللتعرف أكثر على الموضوع نفرض أن لدينا مجتمعا يتكون من N وحدة وهي $X_1,\dots,X_N$

هذه الوحدات إما أن تملك خاصية معينة وإما أن لا تملكها. نفرض أن الذي يملك خاصية معينة سيعطى الرقم واحد وان الذي لا يملك هذه الخاصية سيعطى الرقم صفر، وبذلك تصبح البيانات الموجودة عندنا أما صفر أو واحد، وهذا ما يحدث كثيرا في الأبحاث الخاصة بسوق السلعة وفي ألأبحاث الخاصة باستطلاع الرأي والأبحاث الخاصة بمراقبة الجودة، حيث تكون البيانات المتاحة في صورة نعم أو لا. فلو أعطينا الرقم واحد للشخص الذي أجاب بـ "نعم" فإننا سنعطي الرقم صفر للشخص الذي أجاب بـ "لا"، وبذلك يمكن حساب متوسط الإجابات والانحراف المعياري لها. فمتوسط الإجابات هو:

$$P = \frac{\sum_i X_i}{N}$$

ويطلق على هذا المتوسط أسم النسبة وهو يساوي عدد الذين يملكون خاصية معينة مقسوما على حجم المجتمع.

نظرية (5): إذا سحبت عينة عشوائية حجمها n من مجتمع يخضع لتوزيع ذي الحدين $b(np, npq)$*، فان المتغير العشوائي $\hat{p}$ يخضع تقريبا لتوزيع طبيعي بوسط حسابي قدره p وتباين قدره $\frac{p(1-p)}{n}$ عندما تكون n كبيرة. أي أن المتغير:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

يخضع لتوزيع طبيعي معياري $N(0, 1)$.

- **مثال (9)**: نفرض أن بحثا أجري على 5 أشخاص لمعرفة استخدامهم لنوع معين من الصابون ، وكانت إجاباتهم كالتالي:

لا ، نعم ، نعم ، لا ، نعم

---

* حيث أن الوسط الحسالي لتوزيع ذي الحدين هو (np) والتباين هو (npq).

والمطلوب: (1) حساب النسبة في المجتمع p للذين قالوا نعم.

(2) حساب تباين المجتمع σ_p^2 .

(3) التوزيع العيني للنسبة إذا سحبت كل العينات من الحجم 2 بإرجاع.

(4) العلاقة بين النسبة في المجتمع p وتوقع النسبة $E(\hat{p})$.

(5) العلاقة بين النسبة في المجتمع وبين تباين النسبة $Var(\hat{p})$.

الحل:

(1) يمكن تحويل البيانات أعلاه إلى الآتي: 0 ,1 , 1 , 0 , 1 ومنها يمكن حساب النسبة في المجتمع p كما يلي :

$$P = \frac{\sum_i X_i}{N} = \frac{3}{5} = 0.6$$

(2) لحساب تباين المجتمع σ_p^2 كمعلمة ثانية فإننا نستعمل الصيغة المعروفة التالية:

$$\sigma_p^2 = \frac{1}{N}(\sum X^2 - \frac{(\sum X)^2}{N})$$

وحيث أن البيانات والحالة هذه معطاة في صورة "صفر ، واحد" فإننا نجد أن: $\sum X = Np$ وكذلك : $\sum X^2 = \sum X = Np$ ، وبذلك تصبح صيغة تباين المجتمع كما يلي:

$$\sigma_p^2 = \frac{1}{N}(Np - \frac{(Np)^2}{N})$$
$$= p - p^2 = p(1-p)$$

وكقاعدة عامة فان التباين في حالة كهذه يساوي:

σ_p^2 = (نسبة الذين يملكون خاصية معينة)(نسبة الذين لا يملكون هذه الخاصية)

σ_p^2 = (0.6)(0.4) = 0.24

(3) التوزيع العيني للنسبة: إذا سحبنا عينة مكونة من مفردتين فان عدد العينات هو:

$N^n = 5^2 = 25$ ، تكون بياناتها كالآتي:

لا لا ، لا نعم، لا نعم، لا لا، لا نعم

نعم لا، نعم نعم، نعم نعم، نعم لا، نعم نعم

نعم لا، نعم نعم، نعم نعم، نعم لا، نعم نعم

لا لا، لا نعم، لا نعم، لا لا، لا نعم

نعم لا ، نعم نعم، نعم نعم، نعم لا، نعم نعم

من المعلوم أن احتمال سحب كل عينة من هذه العينات 1/25 وأن نسبة الذين يستخدمون ذلك النوع المعين من الصابون في العينات كالتالي:

0.5 0 0.5 0.5 0

1 0.5 1 1 0.5

1 0.5 1 1 0.5

0.5 0 0.5 0.5 0

1 0.5 1 1 0.5

وبذلك يمكن إيجاد التوزيع العيني للنسبة كما في الجدول أدناه.

جدول رقم (8)

التوزيع العيني للنسبة $\hat{p}$ في حالة السحب بالإرجاع

$\hat{p}$	0	0.5	1
$p(\hat{p})$	4/25	12/25	9/25

واضح جدا أن $p(\hat{p})$ توزيع احتمالي له المتوسط والتباين الآتيين:

$$(a)\ \ E(\hat{p}) = \sum \hat{p}\ p(\hat{p})$$

$$= 0(\frac{4}{25}) + 0.5(\frac{12}{25}) + 1(\frac{9}{25}) = 0.6$$

وهذا يعني أن $E(\hat{p}) = p$.

$$(b)\ \ Var(\hat{p}) = \sum (\hat{p} - E(\hat{p}))^2\ \ p(\hat{p})$$

والجدول التالي يبين خطوات حساب التباين.

جدول (رقم 9)

حساب تباين التوزيع العيني للنسبة

$\hat{p}$	$p(\hat{p})$	$\hat{p} - E(\hat{p})$	$(\hat{p} - E(\hat{p}))^2\ p(\hat{p})$
0	4/25	-0.6	0.0576
0.5	12/25	-0.1	0.0098
1	9/25	0.4	0.0576
المجموع			**0.12**

أي أن $Var(\hat{p}) = 0.12$. ولقد تم إثبات أن هناك علاقة بين تباين التوزيع العيني للنسبة وتباين المجتمع. هذه العلاقة هي:

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

وبالتعويض في هذه العلاقة نجد أن:

$$Var(\hat{p}) = \frac{(0.6)(0.4)}{2} = 0.12$$

من هنا يتضح أن التوزيع العيني للنسبة له الخصائص التالية:

أ) توقع النسبة في التوزيع العيني يعطي النسبة في المجتمع، أي أن: $E(\hat{p}) = p$.

ب) العلاقة بين تباين النسبة $Var(\hat{p})$ وبين النسبة في المجتمع p تكون على الشكل الآتي:

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

ج) باستخدام نظرية الحد المركزية نجد أن النسبة في العينة تتوزع توزيعا معتدلا عياريا بمتوسط p وانحراف معياري $\sqrt{\frac{p(1-p)}{n}}$ إذا كانت n كبيرة، أي أن:

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0.1)$$

إذا كان السحب بإرجاع حيث يمكن استخدام جميع خصائص المنحنى المعتدل.

د) إذا كان السحب بدون إرجاع فإنه باستخدام نظرية النهاية المركزية وجد أن النسبة في العينة تتبع التوزيع المعتدل المعياري بمتوسط يساوي p وتباين قدره $\frac{p(1-p)}{n}.\frac{N-n}{N-1}$ إذا كانت n كبيرة، أي أن المتغير Z:

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}.\frac{N-n}{N-1}}} \sim N(0.1)$$

ويمكن استخدام جداول المنحنى المعتدل المعياري لحساب الاحتمالات مع ملاحظة أن تباين النسبة في حالة السحب بدون إرجاع يكون دائما أقل من التباين للنسبة في حالة الإرجاع. وعندما تكون n كبيرة فإن التباين في الحالتين يكون متساويا.

(7.1) التوزيع العيني للفرق بين نسبتين

نظرية (6)

إذا سحبت عينتان مستقلتان حجم العينة الأولى n وحجم العينة الثانية m من مجتمعين يخضع الأول منهما إلى توزيع ذي الحدين $b(np_1, np_1q_1)$ والثاني إلى توزيع ذي حدين $b(mp_2, mp_2q_2)$، فإن التوزيع العيني لفرق النسبة $\hat{p}_1 - \hat{p}_2$ يخضع تقريبا للتوزيع الطبيعي بوسط حسابي يساوي $p_1 - p_2$ وتباين

$\frac{p_1(1-p_1)}{n}+\frac{p_2(1-p_2)}{m}$ وذلك عندما تكون *n, m* كبيرتين وبالتالي فإن المتغير:

$$Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n}+\frac{p_2(1-p_2)}{m}}} \sim N(0.1)$$

(8.1) التوزيع العيني للتباين (*Sampling Distribution of* S^2)

نظرية (7)

إذا كانت $X_1,\dots,X_n$ عينة عشوائية تباينها S^2 مأخوذة من توزيع طبيعي $N(\mu,\sigma^2)$ وأن σ^2 معلومة فإن المتغير χ^2 :

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}$$

يخضع لتوزيع مربع كاي بدرجات حرية *n*-1.

- **مثال (10):** مجتمع مكون من ثلاث مفردات (*N*=3) حيث: $X_1=1,\ X_2=2,\ X_3=3$. فإذا سحبت كل العينات التي من الحجم 2 بالإرجاع، أوجد التوزيع العيني للتباين S^2 ثم احسب متوسط هذا التوزيع واذكر العلاقة بينه وبين تباين المجتمع.

الحل:

من المعلوم أن عدد العينات الممكنة يساوي $N^n=3^2=9$ وان احتمال سحب أي عينة من هذه العينات يساوي 1/9. والجدول التالي يبين عدد العينات والوسط الحسابي لكل عينة بالإضافة إلى حساب التباين من كل عينة.

جدول (10)

العينات	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
$\bar{X}$	1	1.5	2	1.5	2	2.5	2	2.5	3
$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$	0	0.5	2	0.5	0	0.5	2	0.5	0

ويكون التوزيع العيني للتباين كما يلي.

جدول (رقم 11)

التوزيع العيني للتباين S^2

S^2	$p(S^2)$
0	3/9
0.5	4/9
2.0	2/9
المجموع	**1**

ومن الجدول السابق نجد أن متوسط التوزيع العيني يكون كما يلي:

$$E(S^2) = \sum S^2 \; p(S^2)$$
$$= 0(\frac{3}{9}) + 0.5(\frac{4}{9}) + 2(\frac{2}{9}) = \frac{2}{3}$$

وإذا ما حسبنا تباين المجتمع المسحوبة منه العينات فهو:

$$\sigma^2 = \frac{1}{N} \sum (X - \mu)^2$$
$$= \frac{1}{3}\left[(1-2)^2 + (2-2)^2 + (3-2)^2\right] = \frac{2}{3}$$

مما يؤكد وجود علاقة بين متوسط التوزيع العيني للتباين وبين تباين المجتمع في حالة السحب بالإرجاع، أي أن : $E(S^2) = \sigma^2 = \frac{2}{3}$.

غير أن هذه العلاقة غير قابلة للتحقق عندما يكون السحب بدون إرجاع، أي أن: $E(S^2) \neq \sigma^2$.

(9.1) توزيع النسبة بين تبايني عينتين

Sampling Distribution of S_1^2 / S_2^2

نظرية (8)

إذا كان S_1^2 هو تباين العينة $X_1,\ldots,X_n$ والتي تخضع لتوزيع طبيعي $N(\mu_1,\sigma_1^2)$ وكان S_2^2 هو تباين العينة $Y_1,\ldots,Y_m$ والخاضعة كذلك لتوزيع طبيعي $N(\mu_2,\sigma_2^2)$ والمستقل عن التوزيع الأول، فإن المتغير:

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}$$

يخضع لتوزيع *F* على درجات الحرية (*n*-1)، (*m*-1) .

- **مثال (11)**: عينتان عشوائيتان مستقلتان الأولى حجمها 25 مفردة مأخوذة من توزيع طبيعي بوسط μ_1 وتباين σ_1^2 ، والثانية حجمها 13 مفردة مأخوذة من مجتمع آخر طبيعي يتوزع بوسط μ_2 وتباين σ_2^2 . والمطلوب إيجاد العدد *c* بحيث أن:

$$p(\frac{S_1^2}{S_2^2} \geq c) = 0.01$$

الحل: لأن $\frac{S_1^2}{S_2^2}$ يخضع لتوزيع *F* بدرجات حرية (1-25)، (1-13) فإنه من جدول *F* مباشرة نجد أن *c = 3.78*.

(10.1) ملخص لتوزيعات المعاينة *Summary of Sampling Dist*

التوزيع المناسب للحل	وصف الحالة	المتغير العشوائي
$Z=\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma}$ ، (بالإرجاع) $Z=\frac{(\bar{X}-\mu)}{\sqrt{\frac{\sigma^2}{n}\frac{N-n}{N-1}}}$ ، (بدون إرجاع)	التوزيع طبيعي، التباين σ^2 معلوم، حجم العينة n.	$\bar{X}$
$t_{n-1}=\frac{(\bar{X}-\mu)\sqrt{n}}{S}$	التوزيع طبيعي، التباين σ^2 غير معلوم، حجم العينة n صغير.	$\bar{X}$
$Z=\frac{(\bar{X}-\mu)\sqrt{n}}{S}$	التوزيع طبيعي، التباين σ^2 غير معلوم، حجم العينة كبير.	$\bar{X}$
$Z\cong\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma}$	التوزيع غير طبيعي، التباين σ^2 معلوم، $n>30$.	$\bar{X}$
$\chi^2=\frac{(n-1)S^2}{\sigma^2}$	التوزيع طبيعي، التباين σ^2 معلوم، حجم العينة n.	S^2
$\chi^2\cong\frac{(n-1)S^2}{\sigma^2}$	التوزيع غير طبيعي، التباين σ^2 معلوم، $n>30$.	S^2
$F_{n-1,m-1}=\frac{S_1^2}{S_2^2}$	التوزيع طبيعي، حجوم العينات n، m .	$\frac{S_1^2}{S_2^2}$

المتغير العشوائي	وصف الحالة	التوزيع المناسب للحل
$\frac{S_1^2}{S_2^2}$	التوزيع غير طبيعي، $n+m$ كبيرة.	$F_{n-1,m-1} \cong \frac{S_1^2}{S_2^2}$
$\hat{p}$	التوزيع غير طبيعي، حجم العينة كبير.	$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

ملاحظة: عندما يكون التوزيع للمجتمع الأصلي غير طبيعي وحجم العينة صغيرا فلا يوجد هناك توزيع مناسب للحل.

تمـــارين

(1) افترض أن مجتمعا يتألف من القيم الآتية: 4 ,2 ، المطلوب:

(أ) إيجاد متوسط وتباين المجتمع.

(ب) رسم المدرج التكراري للتوزيع العيني للمتوسط إذا سحبت جميع العينات الممكنة ذات الحجم 4 بالإرجاع.

(ج) من خلال التوزيع العيني للمتوسط، اثبت أن:

$$Var(\overline{X}) = \sigma^2 / n, \quad E(\overline{X}) = \mu .$$

(2) من مجتمع يتكون من 4 مفردات وهي: 5 ,4 ,3 ,2 المطلوب:

(أ) حساب تباين المجتمع.

(ب) إيجاد التوزيع العيني للتباين عندما يكون حجم العينات 3 والسحب بإرجاع، ثم اثبت أن $E(S^2) = \sigma^2$.

(ج) إيجاد التوزيع العيني للتباين عندما يكون حجم العينات 3 والسحب بدون إرجاع، ثم اثبت أن : $E(S^2) \neq \sigma^2$.

(3) من مجتمع يتكون من 5 طلاب وجد أن ثلاثة منهم يدخنون، فإذا أعطينا الرقم واحد للطالب المدخن فان مفردات المجتمع هي:

$X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0$. والمطلوب:

(أ) سحب كل العينات من الحجم 2 بدون إرجاع.

(ب) اثبت أن توقع النسبة في التوزيع العيني يساوي النسبة في المجتمع.

(ج) إذا كان السحب بدون إرجاع، أوجد التوزيع العيني للنسبة وأثبت أن: $E(\hat{p}) = p$.

(د) قارن بين تشتتي التوزيع العيني للنسبة في حالة السحب بإرجاع وبدون إرجاع.

(4) تكلم عن نظرية النهاية المركزية من حيث هدفها والشروط التي يجب استيفائها قبل تطبيق هذه النظرية.

(5) افرض أن لدينا المجتمعين A, B الآتيين:

المجتمع	n	μ	σ
A	49	30	3
B	64	35	6

والمطلوب:

(أ) إيجاد متوسط الفرق بين الوسطين الحسابين.

(ب) حساب تباين متوسط الفرق بين الوسطين الحسابين.

(6) اذكر توزيع المعاينة لكل من: $\hat{p}, S^2, \bar{X}-\bar{Y}, \hat{p}_1-\hat{p}_2$.

(7) أخذت عينة حجمها 100 من مجتمع طبيعي بوسط حسابي قدره 10 وتباين قدره 25. احسب الوسط الحسابي والتباين لمتوسط العينة.

(8) استعمل جدول الأرقام العشوائية المرفق لسحب عينة عشوائية بسيطة حجمها 10 من بين أفراد مؤسسة حكومية مكونة من 100 موظفا.

(9) لماذا تفضل دراسة جزء من المجتمع بدلا من دراسة المجتمع ككل؟

(10) كانت نسبة الطالبات في جامعة ما 0.45 وكانت نسبتهن في جامعة أخرى 0.65 فإذا سحبت عينة من طلبة الجامعة الأولى حجمها 100 وأخرى من الجامعة الثانية حجمها 200 ، المطلوب حساب الاحتمال الآتي: $p(|\hat{p}_1-\hat{p}_2| \leq 0.15$ حيث $\hat{p}_1$, $\hat{p}_2$ تمثلان نسب الإناث في الجامعتين الأولى والثانية على التوالي.

الفصل الثاني

الاستدلال الإحصائي

(1.2) مقدمة.

(2.2) خصائص التقدير الجيد.

(3.2) تقدير معالم المجتمع.

(4.2) خطأ المعاينة والدقة.

(5.2) تقدير حجم العينة.

الاستدلال الإحصائي

Statistical Inference

(1.2) مقدمة (*Introduction*)

سبق وأن تعرفنا على التوزيعات الاحتمالية، كما أننا درسنا خصائص التوزيعات العينية وكما هو معروف بأن الاحصاء يقسم الى الاحصاء الوصفي والاحصاء الاستدلالي وهذا كله ينقلنا إلى موضوع أساس وهو تقدير معالم المجتمع. وتعتمد دراسة المجتمعات على أخذ كل مفردات المجتمع للتعرف على خصائصه وحساب معالمه. والمقصود في المجتمع في الدراسات الإحصائية هو مجموعة المفردات التي تتصف بواحدة أو أكثر من الصفات المشتركة، كمجتمع أطوال الطلاب في كلية مزايا مثلا. والقيمة التي تحسب من توزيع المجتمع لدراسة خصائصه تسمى معلمة (*Parameter*). فالوسط الحسابي للمجتمع هو معلمة لهذا المجتمع ويرمز له بالرمز μ ، كما أن انحرافه المعياري هو معلمة لهذا المجتمع أيضا ويرمز له بالرمز σ، وأخيرا فإن نسبة ظاهرة ما في المجتمع هو معلمة لهذا المجتمع والتي نرمز لها بالرمز p...الخ. وبصفة عامة فالمعالم هي مقادير ثابتة للمجتمع الواحد، ولكنها تتغير من مجتمع إلى آخر.

غير أن دراسة المجتمعات للتعرف على القيم الحقيقية لمعالمها عن طريق أخذ كل مفردات المجتمع (المسح الشامل) له من العيوب التي تجعلنا نفضل العينة عليه، فعملية المسح هذه تحتاج إلى وقت طويل واعتمادات مالية كبيرة وجهد ليس بقليل إلى جانب تعذر دراسة كل مفردات المجتمع أصلا في بعض الدراسات مثل مجتمع الاسماك في

بحيرة ما أو مخزون العراق من البترول. ولكل هذه الاعتبارات السابقة فإنه يكون من المفيد من الناحية العملية ألا نفحص المجتمع فردا فردا ولكننا نكتفي بفحص جزء منه والذي نطلق عليه اسم العينة (*Sample*). وعند دراسة العينات فإن أي قيمة تحسب من العينة المختارة مثل الوسط الحسابي $\overline{X}$ أو الانحراف المعياري S تسمى إحصاءة (*Statistic*) علما بأن قيمة كل إحصائية تختلف من عينة إلى أخرى. غير أننا تعلمنا من الباب السابق أن هناك عددا كبيرا من العينات يمكن سحبها من المجتمع، والسؤال أي من هذه العينات تصلح أو تستخدم لإيجاد القيمة المقدرة لمعلمة المجتمع المجهولة؟ ثم هل يمكن تقدير معلمة المجتمع من عينة واحدة؟ وإذا فرض أننا استخدمنا عينة واحدة لإيجاد القيمة المقدرة لمعلمة المجتمع فهل هذه القيمة جيدة أم لا؟ بمعنى أننا نريد أن ندرس خصائص التقدير الجيد وذلك بأن نضع معاييرا للتقدير الجيد. وفي كل الأحوال فإن كلمة تقدير تعني أننا سوف نحصل على قيمة تقديرية لمعلمة المجتمع. وبالتالي يتحتم علينا تقدير كمية الخطأ التي وقعنا فيها نتيجة لهذا التقدير والتأكد ومن أنها تقع في حدود معينة مقبولة والاستدلال الإحصائي هو الذي يساعدنا على إيجاد الطرق المختلفة لتقدير معالم المجتمع المجهولة ويساعدنا كذلك على دراسة خصائص هذا التقدير والتعرف على كمية الأخطاء التي سنتعرض لها.

(2.2) خصائص التقدير الجيد

Properties of a good Estimator

إذا كان هدفنا في الاستدلال الإحصائي هو الوصول إلى تقدير جيد لمعلمة المجتمع المجهولة، فإنه وقبل التعرض لطرق التقدير المختلفة يتحتم أن نضع معاييرا لمعنى كلمة (جيد) من وجهة النظر الإحصائية. فمتوسط المجتمع مثلا يمكن تقديره عن طريق متوسط العينة أو عن طريق وسيط العينة. ولكن أي من هذين التقديرين هو الأحسن في التقدير؟ وفي الحقيقة نحن نعني بالتقدير الأحسن أو التقدير الجيد أنه التقدير الذي يكون توزيعه العيني يتركز حول المعلمة المجهولة. ومن هنا يبدأ الدور

الذي يلعبه التوزيع العيني في الاستدلال الإحصائي. وخصائص التقدير الجيد الذي سنتعرف عليها تجيب بمجموعها على السؤال الذي يقول: هل أن التوزيع العيني للتقدير متمركز حول معلمة المجتمع أم لا ؟ **وهذه الخصائص هي:**

أولا: عدم التحيز (*Unbiasedness*)

إذا كان لدينا مجتمعا مكونا من N وحدة متوسطه μ وهو المعلمة المجهولة والمراد تقديرها. نفرض أننا سحبنا عينة من الحجم n بطريقة عشوائية سواء كان السحب بإرجاع أو بدون إرجاع ثم حسبنا متوسط العينة $\bar{X}$ ، فإن كان متوسط التوزيع العيني للمتوسط $E(\bar{X})$ يساوي متوسط المجتمع μ فإننا والحالة هذه نقول إن $\bar{X}$ هو تقدير غير متحيز لمتوسط المجتمع μ ، وهذا لا يتوفر في الإحصائيات الأخرى كالوسيط والمنوال مثلا. ولو استطعنا أن نثبت ذلك رياضيا فإنه يكفي أن نسحب عينة واحدة فقط من الحجم n ونحسب متوسط هذه العينة حيث يمكن اعتباره تقديرا غير متحيز لمتوسط المجتمع.

ولإثبات أن: $E(\bar{X}) = \mu$

فإن:

$$E(\bar{X}) = E(\frac{\sum X_i}{n})$$
$$= \frac{1}{n} E(X_1 + \cdots + X_n)$$

وباستخدام خواص التوقع فان:

$$E(\bar{X}) = \frac{1}{n}\left[E(X_1) + \cdots + E(X_n)\right]$$

ولأن $E(X_i) = \mu$ فان:

$$E(\bar{X}) = \frac{1}{n}(\mu + \cdots + \mu)$$
$$= \frac{1}{n}(n\mu) = \mu$$

كما أن نسبة العينة $\hat{p}$ (والتي تمثل نسبة امتلاك صفة معينة) تعتبر إحصائية غير متحيزة للنسبة في المجتمع p. ويمكن إثبات ذلك رياضيا كما يلي:

$$\hat{p} = \frac{x}{n}$$

$$E(\hat{p}) = E(\frac{x}{n}) = \frac{1}{n}E(x)$$

ولأن المتغير X يتوزع توزيعا ذات حدين بوسط يساوي np وتباين قدره npq فان:

$$E(\hat{p}) = \frac{1}{n}(np) = p$$

وكذلك الحال بالنسبة إلى تباين العينة S^2 فهو تقدير غير متحيز لتباين المجتمع σ^2 إذا ما تم استخدام الصيغة:

$$S^2 = \frac{1}{n-1}\sum(X-\overline{X})^2$$

ويمكن البرهنة على ذلك رياضيا كما يلي:

$$S^2 = \frac{1}{n-1}\sum(X-\overline{X})^2$$

$$S^2 = \frac{1}{n-1}\sum((X-\mu)-(\overline{X}-\mu))^2$$

وذلك بإضافة وطرح μ .

$$S^2 = \frac{1}{n-1}\sum\left[(X-\mu)^2 - 2(X-\mu)(\overline{X}-\mu) + (\overline{X}-\mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum(X-\mu)^2 - 2n(\overline{X}-\mu)^2 + n(\overline{X}-\mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum(X-\mu)^2 - n(\overline{X}-\mu)^2\right]$$

لأن $\sum(X-\mu)(\bar{X}-\mu)=n(\bar{X}-\mu)^2$ ، وبأخذ توقع الطرفين:

$$E(S^2)=\frac{1}{n-1}E\left[\sum(X-\mu)^2-n(\bar{X}-\mu)^2\right]$$

$$=\frac{1}{n-1}\left[\sum E(X-\mu)^2-nE(\bar{X}-\mu)^2\right]$$

$$=\frac{1}{n-1}\left[\sum\sigma^2-n(\frac{\sigma^2}{n})\right]=\frac{1}{n-1}(n\sigma^2-\sigma^2)$$

لأن: $E(X-\mu)^2=\sigma^2$ وان: $E(\bar{X}-\mu)^2=\frac{\sigma^2}{n}$ ومنها نجد أن:

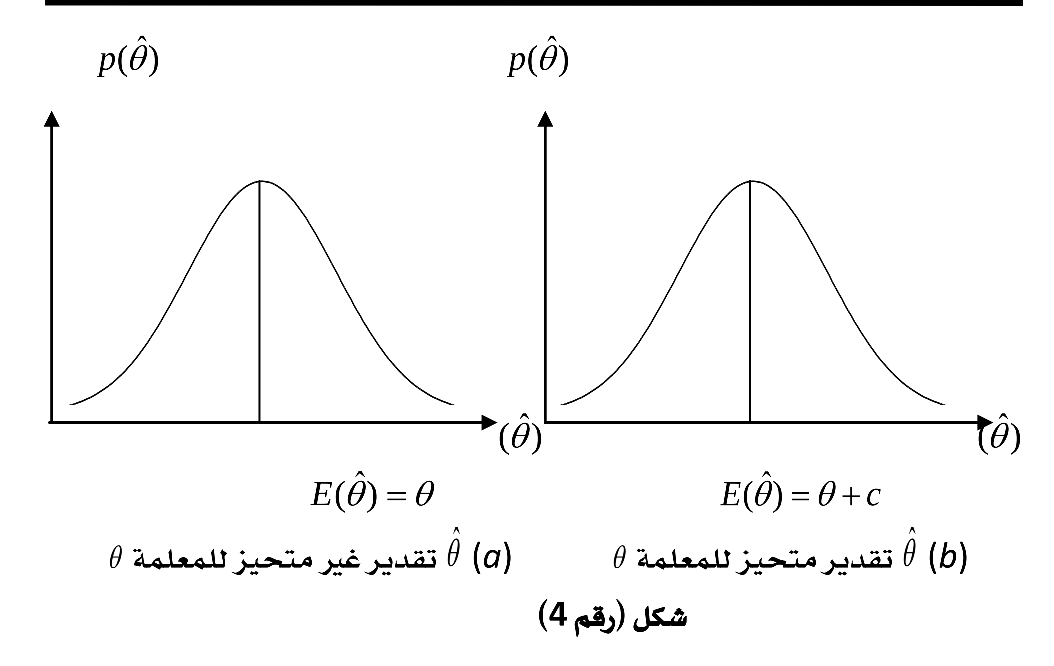
$$E(S^2)=\frac{(n-1)}{(n-1)}\sigma^2=\sigma^2$$

ومن هنا يتضح أنه لكي يكون التقدير غير متحيز فإنه يجب أن يكون التوزيع العيني للتقدير متمركزا حول معلمة المجتمع المجهولة أيا كان اسمها. أما إذا كان التوزيع العيني للتقدير لا يتمركز حول معلمة المجهولة للمجتمع فإننا نقول إن التقدير المحسوب من العينة تقدير متحيز لمعلمة المجتمع المجهولة، كما هو الحال بالنسبة لتباين العينة S^2 إذا ما تم استعمال $S^2=\frac{1}{n}\sum(X-\bar{X})^2$ ، فهو تقدير متحيز لتباين المجتمع σ^2 ، بمعنى أن $E(S^2)\neq\sigma^2$.

قاعدة عامة: إذا كانت θ هي معلمة المجتمع المجهولة وكانت $\hat{\theta}$ هي التقدير المحسوب من العينة، فإننا نلاحظ الحالتين الآتيتين:

(1) إذا كان $E(\hat{\theta})=\theta$ ، فإن $\hat{\theta}$ تعتبر تقدير غير متحيز للمعلمة كما هو الحال في (a) من الشكل رقم (4).

(2) إذا كان $E(\hat{\theta})=\theta+c$ ، حيث c مقدار ثابت فإن $\hat{\theta}$ تعتبر تقديرا متحيزا للمعلمة θ لأن التوزيع العيني للمتغير $\hat{\theta}$ يتمركز حول $\theta+c$ وليس حول θ وحدها كما في (b) من الشكل التالي.



(a) $\hat{\theta}$ تقدير غير متحيز للمعلمة θ (b) $\hat{\theta}$ تقدير متحيز للمعلمة θ

شكل (رقم 4)

ويسمى الحد c بحد التحيز *(Bias term)* وقد يكون هذا الحد موجبا أو سالبا، فإن كان موجبا فهذا يعني أن: $E(\hat{\theta}) > \theta$ وهنا يقال أن تحيز $\hat{\theta}$ تحيز موجب *(Positively biased)* أما إذا كانت c كمية سالبة فهذا يعني أن $E(\hat{\theta}) < \theta$ وهنا يقال أن تحيز $\hat{\theta}$ تحيز سالب *(Negatively biased)*).

- **مثال (12)**: من مجتمع يتكون من 3 وحدات حيث: $X_1 = 1, X_2 = 2, X_3 = 3$ سحبت عينات من الحجم 2. أوجد التوزيع العيني للمتوسط واثبت أن متوسط العينة تقدير غير متحيز لمتوسط المجتمع سواء كان السحب بإرجاع أو بدون إرجاع.

الحل: (أ) إذا كان السحب بإرجاع فان عدد العينات هو 9 وبياناتها كما يلي:

1,1	1,2	1,3
2,1	2,2	2,3
3,1	3,2	3,3

وان احتمال الحصول على أي عينة هو 1/9 ، والجدول التالي يبين التوزيع العيني للمتوسط.

جدول (رقم 12)

$\bar{X}$	1	1.5	2	2.5	3
$p(\bar{X})$	1/9	2/9	3/9	2/9	1/9

ولإثبات أن متوسط أي عينة من العينات السابقة يعتبر تقديرا غير متحيز لمتوسط المجتمع نجري الآتي:

$$E(\bar{X}) = \sum \bar{X}\ p(\bar{X})$$
$$= 1(1/9) + 1.5(2/9) + 2(3/9) + 2.5(2/9) + 3(1/9) = 2$$

وحيث أن متوسط المجتمع μ يساوي 2 فان $E(\bar{X}) = \mu$ ، وبذلك نستنتج أن متوسط العينة هو تقدير غير متحيز لمتوسط المجتمع.

(ب) إذا كان السحب بدون إرجاع فان عدد العينات هو 3 وهي: (2,3), (1,3), (1,2)
وان احتمال الحصول على أي عينة هو 1/3. والجدول التالي يبين التوزيع العيني وهذه الحالة.

جدول (رقم 13)

$\bar{X}$	1.5	2	2.5
$p(\bar{X})$	1/3	1/3	1/3

وبالتالي فان:

$$E(\bar{X}) = \sum \bar{X}\ p(\bar{X})$$
$$= 1.5(1/3) + 2(1/3) + 2.5(1/3) = 2$$

وحيث أن $\mu = 2$ فان متوسط العينة في هذه الحالة يكون هو الآخر تقديرا غير متحيز لمتوسط المجتمع.

مما تقدم عرضه فانه يمكن سحب عينة واحدة فقط من مفردتين سواء كان السحب بإرجاع أو بدون إرجاع ثم نقوم بحساب متوسط أي عينة ليكون تقديرا غير متحيز لمتوسط المجتمع. وهذا معناه أن التوزيع العيني للمتوسط في حالة الإرجاع وعدم الإرجاع يتمركز حول معلمة المجتمع المجهولة.

- **مثال (13):** في مثالنا السابق إذا كنا نهتم بنسبة الأرقام الفردية في العينة. المطلوب إيجاد التوزيع العيني للنسبة في حالة الإرجاع وعدم الإرجاع ثم إثبات أن النسبة في العينة تقدير غير متحيز للنسبة في المجتمع.

الحل:

(أ) في حالة الإرجاع فان التوزيع العيني للنسبة يمكن إيجاده باستخدام العينات المذكورة في المثال السابق وتمثيله في الجدول أدناه.

جدول (رقم 14)

$\hat{p}$	0	0.5	1
$p(\hat{p})$	1/9	4/9	4/9

ولإثبات أن النسبة في العينة تقدير غير متحيز للنسبة في المجتمع فإننا نحسب توقع $\hat{p}$ والتي ترمز إلى النسبة في **العينة كما يلي:**

$$E(\hat{p}) = \sum \hat{p}\, p(\hat{p})$$

ولأن

$$= 0(1/9) + 0.5(4/9) + 1(4/9) = 2/3$$

نسبة

الأرقام الفردية في المجتمع هي 2/3 وهي معلمة المجتمع المجهولة p، فهذا معناه أن: $E(\hat{p}) = p = 2/3$ ، مما يدل على أن التوزيع العيني للنسبة في حالة الإرجاع يتمركز حول معلمة المجتمع.

(ب) في حالة عدم الإرجاع فانه يمكن إيجاد التوزيع العيني للنسبة في الجدول التالي حيث أن عدد العينات 3 وهي: (0,1) ,(1,1) ,(1,0) .

جدول (رقم 15)

$\hat{p}$	0.5	1
$p(\hat{p})$	2/3	1/3

ومن الجدول نجد أن توقع النسبة هو:

$$E(\hat{p}) = \sum \hat{p}\, p(\hat{p})$$
$$= 0.5(2/3) + 1(1/3) = 2/3$$

وبالتالي فان: $E(\hat{p}) = p = 2/3$. ومن هنا يمكن القول أن النسبة في العينة هي تقدير غير متحيز للنسبة في المجتمع،أي أن التوزيع العيني لنسبة الأرقام الفردية في العينة يتمركز حول نسبة الأرقام الفردية في المجتمع سواء كان السحب بإرجاع أو بدون إرجاع.

- **مثال (14)**: عودة للمثال (12)، أوجد التوزيع العيني للتباين ثم اثبت أن التباين في العينة S^2 هو تقدير غير متحيز للتباين في المجتمع σ^2 إذا كان السحب بالإرجاع وهو تقدير متحيز لـ σ^2 إذا كان السحب بدون إرجاع.

الحل:

(أ) إذا كان السحب بالإرجاع فإن عدد العينات هو 9 كما سبق وأن ذكرنا. وبإيجاد التباين في كل عينة باستخدام الصيغة $S^2 = \frac{1}{n-1}\sum(X-\bar{X})^2$ فإنه يمكن إيجاد التوزيع العيني للتباين في حالة الإرجاع كما في الجدول الآتي.

جدول (رقم 16)

S^2	0	0.5	2
$p(S^2)$	3/9	4/9	2/9

ولإثبات أن التباين في العينة S^2 تقدير غير متحيز للتباين في المجتمع σ^2 فإن:

$$E(S^2) = \sum S^2 \; p(S^2)$$
$$= 0(3/9) + 0.5(4/9) + 2(2/9) = 2/3$$

وهذا يساوي σ^2 حيث أن:

$$\sigma^2 = \frac{1}{N}\sum (X-\mu)^2$$
$$= \frac{1}{3}\left[(1-2)^2 + (2-2)^2 + (3-2)^2\right] = 2/3$$

أي أن $E(S^2) = \sigma^2$ ، مما يعني أن التباين في العينة هو تقدير غير متحيز لتباين المجتمع.

(ب) في حالة السحب بدون إرجاع فإن التوزيع العيني للتباين يوضحه الجدول الآتي.

جدول (رقم 17)

S^2	0.5	2
$p(S^2)$	2/3	1/3

ولإثبات أن التباين في العينة S^2 تقدير متحيز لتباين المجتمع σ^2 فإن:

$$E(S^2) = \sum S^2 \; p(S^2)$$
$$= 0.5(2/3) + 2(1/3) = 1$$

وهذا يعني أن $E(S^2) \neq 2/3$ ، وبذلك يكون تباين العينة في حالة عدم الإرجاع تقديرا متحيزا لتباين المجتمع σ^2 طالما أن $E(S^2) \neq \sigma^2$.

ثانيا: الاتساق $(Consistency)$

وهي الخاصية الثانية التي يجب أن تتوفر في التقديرات الجيدة. ويقصد بالاتساق انه يمكن لإحصائية العينة أن تؤول إلى معلمة المجتمع كلما زاد حجم العينة المسحوبة عشوائيا من المجتمع الأصلي وذلك نتيجة لانخفاض درجة التشتت بين القيم الاحتمالية والقيم الحقيقية.

فإذا فرضنا أن لدينا مجتمعا من الحجم N متوسطه μ ، وقد سحبنا منه عينة من الحجم n متوسطها $\overline{X}$ ، فإنه يقال أن $\overline{X}$ تقدير متسق لمتوسط المجتمع μ إذا اقتربت قيمته من قيمة متوسط المجتمع عندما تكبر n كبرا كافيا. ويمكن التعبير عن ذلك رياضيا بالآتي:

$$p(\overline{X} \to \mu) \to 1 \qquad as \quad n \to \infty$$

ويعني هذا التعبير أن احتمال أن يقترب متوسط العينة من متوسط المجتمع هو احتمال مؤكد (يساوي واحد) عندما تكبر n وتؤول إلى مالا نهاية.

ومن الخاصية الأولى والثانية يتضح أن عدم التحيز لا يعني تساوي متوسط العينة مع متوسط المجتمع، فقد يختلف متوسط العينة عن متوسط المجتمع (كما في الأمثلة السابقة) غير أنه بالإمكان اعتباره تقدير غير متحيز. ولكن الاتساق يعني أن المتوسط في العينة يقترب من متوسط المجتمع عندما يكون حجم العينة n كبيرا كبراً كافياً.

قاعدة عامة: نفرض أن معلمة المجتمع المجهولة هي θ وأن التقدير من العينة هو $\hat{\theta}$.

أفرض أن N هو حجم المجتمع وأن حجم العينة هو n، وهنا يقال أن $\hat{\theta}$ تقدير متسق للمعلمة المجهولة θ إذا حدث الآتي:

$$p(\hat{\theta} \to \theta) \to 1 \qquad as \quad n \to \infty$$

وبلغة التوزيعات العينية نقول أن التوزيع العيني للتقدير $\hat{\theta}$ يكون أكثر تمركزا حول معلمة المجتمع المجهولة θ عندما تكبر n.

وهنا يمكن القول أن $\overline{X}$ تقدير متسق للمعلمة μ ويمكن إثبات ذلك رياضيا كما يلي:

$$\overline{X} = \frac{1}{n}\sum X_i$$

$$Var(\overline{X}) = Var(\frac{1}{n}\sum X_i)$$

$$= \frac{1}{n^2}Var(X_1 + \cdots + X_n)$$

ولكن $X_i \sim N(\mu, \sigma^2)$ ، إذن :

$$Var(\overline{X}) = \frac{1}{n^2}(Var(X_1) + \cdots + Var(X_n))$$

$$= \frac{1}{n^2}(\sigma^2 + \cdots + \sigma^2)$$

$$= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

وبالتالي فإن

$$\lim_{n\to\infty} Var(\overline{X}) = 0$$

وهذا هو المقصود بانخفاض درجة التشتت بين القيم الاحتمالية والقيم الحقيقية، الأمر الذي يعني أن متوسط العينة يقترب من متوسط المجتمع عندما يكبر حجم العينة. ومن جانب آخر فإن S^2 تقدير متسق هو الآخر للمعلمة σ^2 ، يمكن إثبات ذلك رياضيا كما يلي:

$$\chi^2 = \frac{vS^2}{\sigma^2}$$

$$Var(\chi^2) = Var(\frac{vS^2}{\sigma^2})$$

$$Var(\chi^2) = (\frac{v}{\sigma^2})^2\, Var(S^2)$$

ولأن $Var(\chi^2)=2v$ ، إذن:

$$Var(S^2)=\frac{2\sigma^4}{v}$$

حيث v تمثل درجات الحرية وهي تساوي (n-1). وبأخذ النهاية لـ $Var(S^2)$ فإن:

$$\lim_{n\to\infty} Var(S^2)=\lim_{n\to\infty}(\frac{2\sigma^4}{n-1})=0$$

وأخيرا يمكن أن نبرهن أن $\hat{p}$ تقدير متسق للمعلمة p وهي نسبة امتلاك الصفة في المجتمع كما يلي:

$$\hat{p}=\frac{x}{n}$$

$$Var(\hat{p})=\frac{1}{n^2}Var(x)$$

ولأن X تتبع توزيع ذي الحدين وأن تباين ذي الحدين هو npq إذن:

$$Var(\hat{p})=\frac{1}{n^2}(npq)=\frac{pq}{n}$$

وبأخذ النهاية لتباين $\hat{p}$ فإن هذه النهاية سوف تؤول للصفر عندنا تكبر n أي أن:

$$\lim_{n\to\infty} Var(\hat{p})=0$$

- **مثال (15):** أفرض أن لدينا مجتمعا حجمه 3000 وحدة، وكان وسطه 80 ووسيطه 85. فإذا سحبت عينة عشوائية من الحجم n فإنه كلما زاد حجم العينة فإن متوسط العينة يقترب من المعلمة المجهولة μ على عكس وسيط العينة الذي يبتعد عن المعلمة المجهولة μ ويقترب من الوسيط الخاص بالمجتمع. وبذلك يمكن القول أن متوسط العينة يعتبر تقديرا متسقا لمتوسط المجتمع في حين أن وسيط العينة يعتبر تقديرا غير متسق له.

ثالثا: الكفاءة (*Efficiency*)

أثبتنا أن الوسط الحسابي للعينة هو تقدير غير متحيز لمتوسط المجتمع في حالة السحب بإرجاع أو بدون إرجاع. ولقد اعتمدنا في إثبات هذه الخاصية على مقياس النزعة المركزية للتوزيع العيني للمتوسط في كلا الحالتين. ولكن لنا أن نسأل أيهما أفضل السحب بإرجاع أم السحب بدون إرجاع؟ من المعروف في الإحصاء الوصفي أن المقارنة بين التوزيعات على أساس مقياس النزعة المركزية فقط غير كاف، فقد يتساوى متوسطا التوزيعين في حين تختلف مقاييس التشتت بينهما. وفي الحقيقة نحن نفضل التوزيع الأقل تشتتا، وهذا ما كنا نعنيه بأن التوزيع العيني للمتوسط يجب أن يكون أكثر تمركزا حول معلمة المجتمع المجهولة.

وبالرجوع إلى جدول (رقم 11) و(رقم 12) حيث التوزيعات العينية للمتوسط في حالتي السحب بإرجاع وبدون إرجاع لوجدنا أن المدى (أكبر قيمة – أصغر قيمة) في التوزيع الأول هو 2، أما المدى في التوزيع الثاني فهو 1، وهذا يعني أن التوزيع العيني في حالة عدم الإرجاع أكثر تمركزا حول معلمة المجتمع المجهولة من التوزيع في حالة الإرجاع، وبذلك نحن نفضل حالة عدم الإرجاع عن حالة الإرجاع ونقول أن التقدير في هذه الحالة أكثر كفاءة من التقدير في حالة الإرجاع. هذا من ناحية، ومن ناحية أخرى فقد وجد أن متوسط العينة هو تقدير غير متحيز وهو تقدير متسق لمتوسط المجتمع إذا كانت المنحنيات متماثلة، وهو الشيء نفسه بالنسبة لوسيط العينة. إذن ما هو التقدير الأفضل هل هو وسط العينة أم وسيطها؟ وهذا يعني إننا يجب أن نبحث في التوزيع العيني للمتوسط والوسيط ونفضل التوزيع صاحب التشتت الأقل والذي سيكون بالإضافة إلى كونه غير متحيز ومتسق سيكون ذو كفاءة عالية *More efficient*).

وبناءا على هذا كله إذا فرضنا أن معلمة المجتمع الذي حجمه N هي μ ، وأن عينة من الحجم n سحبت منه ثم حسب متوسطها فكان $\overline{X}$ وحسب وسيطها فكان M ، فإنه يقال أن الوسط الحسابي أكثر كفاءة من الوسيط إذا حدث الآتي:

$$Var(\overline{X}) < Var(M)$$

قاعدة عامة: إذا كانت معلمة المجتمع المجهولة هي θ وتم حساب $\hat{\theta}_1$ ، $\hat{\theta}_2$ من العينة بحيث أن:

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

فإننا نقول أن التقدير $\hat{\theta}_1$ هو تقدير أكثر كفاءة من التقدير $\hat{\theta}_2$.

- **مثال (16):** في مثال (12)، وفي حالة السحب بإرجاع نجد أن:

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$
$$= \frac{2}{3}.\frac{1}{2} = \frac{1}{3}$$

بينما نجد في حالة السحب بدون إرجاع أن:

$$Var(\overline{X}) = \frac{\sigma^2}{n}.\frac{N-n}{N-1}$$
$$= (\frac{2}{3})(\frac{1}{2})(\frac{3-2}{3-1}) = \frac{1}{6}$$

وهذا يعني أن التباين المتوسط في حالة السحب بدون إرجاع أقل منه في حالة السحب بالإرجاع وهو بذلك يكون أكثر كفاءة.

رابعا: الكفاية (*Sufficiency*)

يقال أن التقدير كاف إذا تم استخدام كل مفردات العينة (أو كل المعلومات المتوفرة عن العينة) في حساب إحصائية العينة. وهذه الخاصية تتوفر للإحصائية $\overline{X}$ دون غيرها من مقاييس النزعة المركزية، إذ نجد أن $\overline{X}$ يعتبر تقديرا كافيا للمعلمة μ حيث يتم استخدام كل مفردات العينة عند حسابه. ولا شك أن طريقة اختيار العينات من المجتمعات الأصلية لها دور أكيد في رفع درجة الكفاية في التقدير.

وبصورة عامة فإنه يتعين على الباحث التأكد من توفر كل أو أكثرية هذه الشروط في التقديرات الجيدة لمعالم المجتمعات. ففي اختيار الوسط الحسابي كتقدير جيد للمعلمة μ علينا أن نتأكد من أن متوسط التوزيع يتمركز أساسا حول الوسط الحسابي للمجتمع الأصلي والذي تم سحب كل العينات منه. كما يجب النظر إلى كل هذه الخصائص دون النظر إلى توفر خاصية واحدة إذا تطلبت الحاجة. بمعنى أننا قد ننظر إلى مجموعة التقديرات الغير متحيزة في حالة توفرها ونبحث عن التقدير ذي التباين الأقل ونفضله، وهذا معناه أننا نظرنا إلى شرطي عدم التحيز والكفاءة في آن واحد عند عملية التفضيل بين التقديرات، وهكذا.

(3.2) تقدير معالم المجتمع

Estimation of Population Parameters

غالبا ما تكون المجتمعات التي نريد معرفة خصائصها مجهولة المعالم كالوسط الحسابي μ والانحراف المعياري σ. وهناك من الأسباب العلمية والاقتصادية ما يحول دون تحديد هذه الخصائص تحديدا مؤكدا ودقيقا. وفي هذه الحالة فإننا نلجأ إلى سحب عينة عشوائية من هذا المجتمع ثم نحسب إحصائيات هذه العينة مثل $\overline{X}, S, \hat{p}$، محاولين تقدير معالم المجتمع الأصلي μ, σ, p من خلال إحصائيات هذه العينة. وهذا ما سنقوم بتفصيله الآن حيث نقسم طرق التقدير إلى نوعين:

(1) التقدير بنقطة *Point estimation*

(2) التقدير بفترة *Interval estimation*

أولا: التقدير بنقطة

وهو أن نعبر عن معلمة المجتمع المجهولة بقيمة واحدة يشارك كل أو بعض مفردات (أو قيم) العينة في تكوينها، بحيث إذا تحققت الشروط الخاصة بالتقدير الجيد يمكن اعتبار هذه القيمة تقديرا جيدا لمعلمة المجتمع. ويعيب هذه الطريقة أنه في كثير من الأحيان تكون قيمة التقدير غير منطبقة على قيمة معلمة

المجتمع، الأمر الذي لا نستطيع معه معرفة مدى دقة التقدير أو مدى بعده عن القيمة الحقيقية المجهولة لمعلمة المجتمع. هذا ويمكن اعتبار أن:

- متوسط العينة $\overline{X}$ تقدير نقطة جيد لمتوسط المجتمع μ .
- الفرق $\overline{X}-\overline{Y}$ تقدير نقطة جيد للفرق $\mu_1-\mu_2$.
- النسبة $\hat{p}$ تقدير نقطة جيد للنسبة p.
- الفرق $\hat{p}_1-\hat{p}_2$ تقدير نقطة جيد للفرق p_1-p_2 .

وهذا يعني إعطاء قيمة محددة لمعلمة المجتمع من خلال إحصائيات العينة. وعند إتباع هذا الأسلوب الإحصائي في التقدير فإنه يجب مراعاة شروط التقدير الجيد ومحاولة توفر أغلبية هذه الشروط أو الخصائص حتى يمكن التوصل إلى أدق التقديرات لمعالم المجتمعات إحصائيا.

فإذا كان متوسط إنتاجية العمال في عينة من 100 عامل هو 170 وحدة يوميا، فإنه يمكن اعتبار هذا المتوسط تقدير غير متحيز ومتسق لمتوسط المجتمع، ونستطيع القول في هذه الحالة أن متوسط المجتمع 170= μ وحدة يوميا.

وهناك طرق أخرى لتقدير معالم المجتمع مثل طريقة المربعات الصغرى وطريقة العزوم وطريقة الإمكان الأعظم وأخيرا طريقة الرسم والتي تعرفنا على أكثرها في كتابنا "طرق التحليل الإحصائي".

ثانيا: التقدير بفترة

الأسباب كثيرة تلك التي دعت الإحصائيين إلى تقدير فترة تقع داخلها معلمة المجتمع المراد الحصول عليها. فالتقدير بنقطة لا يقترن بدرجة معينة من الصحة أو الثقة في النتائج، بالإضافة إلى صعوبة تحقيق الخصائص الواجب توافرها في المقدر الجيد، ناهيك عن أن أخذ العينات ينتج عنه أخطاء من أنواع مختلفة، لذلك كله فمن الأفضل ربط تقدير المعلمة المجهولة بدرجة معينة للثقة في النتائج أو بنسبة معينة للخطأ في التقدير ووضع حدين، حد أعلى (يسمى بالحد

الأعلى للثقة) وحد أدنى (يسمى بالحد الأدنى للثقة)، والفرق بينهما يسمى بمدى الثقة. وفي إطار هذا المدى تقع معلمة المجتمع عند مستوى الثقة المحددة سلفا.

قاعدة عامة: إذا كان لدينا مجتمع يتكون من N وحدة، معلمته المجهولة θ، سحبنا عينة عشوائية منه من الحجم n وحسبنا قيمة $\hat{\theta}$ منها، فإنه باستخدام نظرية النهاية المركزية نجد أن المتغير:

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}}$$

يخضع للتوزيع المعتدل المعياري، وأنه باحتمال $1-\alpha$ نجد أن:

$$P(LL < Z < UL) = 1 - \alpha$$

ومنها:

$$p(\hat{\theta} - Z_{\alpha/2}\sqrt{Var(\hat{\theta})} < \theta < \hat{\theta} + Z_{\alpha/2}\sqrt{Var(\hat{\theta})}) = 1-\alpha$$

حيث:

LL, UL = فترة الثقة — *Confidence Interval*

LL = الحد الأدنى للثقة — *Lower Limit*

UL = الحد الأعلى للثقة — *Upper Limit*

$Z_{\alpha/2}$ = القيمة المعيارية المستخرجة من جداول المنحنى المعتدل المناظرة لاحتمال $\alpha/2$. بمعنى آخر، أن المساحة المحصورة ما بين $Z_{\alpha/2}$ ونهاية التوزيع المعتدل هي $\alpha/2$.

$1-\alpha$ = مستوى الثقة — *Confidence Level*

α = نسبة الخطأ في التقدير.

$\sqrt{Var(\hat{\theta})}$ = الانحراف المعياري لـ $\hat{\theta}$.

وبشكل عام تكون فترة الثقة على الصيغة التالية:

تقدير نقطة $\pm$ (معامل الثقة) (الخطأ المعياري).

والمعنى الطبيعي لما ذكرنا أنه لو سحبنا مائة عينة من الحجم n وحسبنا من كل عينة $\hat{\theta}$ ثم حسبنا حدود الثقة المناظرة للتقدير $\hat{\theta}$ لكل عينة فإننا سوف نجد باحتمال $1-\alpha$ أن معلمة المجتمع تقع في هذه الحدود، وباحتمال قدره α أن هذه المعلمة تقع خارج هذه الحدود.

وسوف تتركز دراستنا هنا على الحالات التي يكون الباحث فيها بصدد مجتمعات مجهولة المعالم تتوزع توزيعا معتدلا أو قريبة من الاعتدال والتعرف على خصائصه من خلال دراستنا لعينة كبيرة أو صغيرة اختيرت مفرداتها عشوائيا.

أولا: تقدير متوسط المجتمع μ

(أ) **إذا كان تباين المجتمع σ^2 معلوما**

لاحظنا أن متوسط العينة $\bar{X}$ هو تقدير غير متحيز لمتوسط المجتمع ، وبالرجوع إلى خصائص التوزيع العيني للمتوسطات الحسابية المسحوبة عشوائيا من المجتمع المراد تقدير المعلمة μ له نجد أن قيمة $\bar{X}$ تقع بالتقريب في الفترة:

$$\left(\bar{X}-\frac{2\sigma}{\sqrt{n}},\ \ \bar{X}+\frac{2\sigma}{\sqrt{n}}\right)$$

وذلك باحتمال قدره 95%. وفي الفترة:

$$\left(\bar{X}-\frac{3\sigma}{\sqrt{n}},\ \ \bar{X}+\frac{3\sigma}{\sqrt{n}}\right)$$

باحتمال قدره 99%. والاحتمال هنا يمثل درجة ثقتنا في التقدير وبالتالي فإن درجة عدم الصحة في التقدير (أي α) للفترة الأولى تساوي: 5% = 95% -1 ودرجة عدم الصحة في التقدير للفترة الثانية تساوي: 1% = 99% -1. أما العددان 3 ,2 فهما يدلان على عدد الانحرافات المعيارية التي تقع ضمنها 95% أو 99% من القيم الممكنة للمتغير $\bar{X}$ على التوالي.

وتسمى هذه الطريقة بتقدير فترة للمعلمة μ بثقة 100% $(1-\alpha)$ عندما يكون تباين المجتمع σ^2 معلوما.

نظرية (9)

إذا كانت $X_1,\dots,X_n$ عينة عشوائية من توزيع طبيعي $N(\mu,\sigma^2)$ ، حيث σ^2 معلومة فإن فترة ثقة 100% $(1-\alpha)$ للمعلمة μ هي:

$$p(\overline{X}-Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})<\mu<\overline{X}+Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}}))=1-\alpha$$

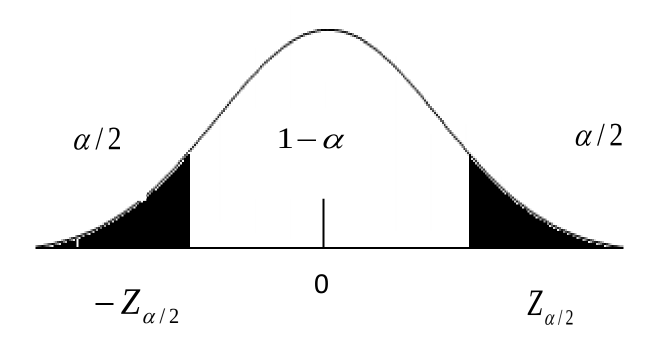
حيث $\overline{X}$ هو الوسط الحسابي للعينة.

البرهان:

إذا كانت $X_1,\dots,X_n$ عينة عشوائية من توزيع طبيعي $N(\mu,\sigma^2)$ ، فإن $\overline{X}$ يخضع لتوزيع طبيعي $N(\mu,\frac{\sigma^2}{n})$. وباستخدام نظرية الحد المركزية فإن:

$$Z=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$$

ومن النظر إلى الشكل التالي:



شكل (رقم 5)

نرى أن احتمال أن تقع Z بين القيمتين $Z_{\alpha/2}$، $Z_{\alpha/2}$ - هو:

$$p(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

وبالتعويض عن قيمة Z بما يساويها ينتج:

$$p(-Z_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\alpha/2}) = 1-\alpha$$

وبضرب حدود المعادلة في $\sigma/\sqrt{n}$ ينتج:

$$p(-Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}}) < \overline{X}-\mu < Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})) = 1-\alpha$$

وبإضافة ($\overline{X}$ -) إلى كل حد من حدود المعادلة نجد أن:

$$p(-\overline{X}-Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}}) < -\mu < -\overline{X}+Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})) = 1-\alpha$$

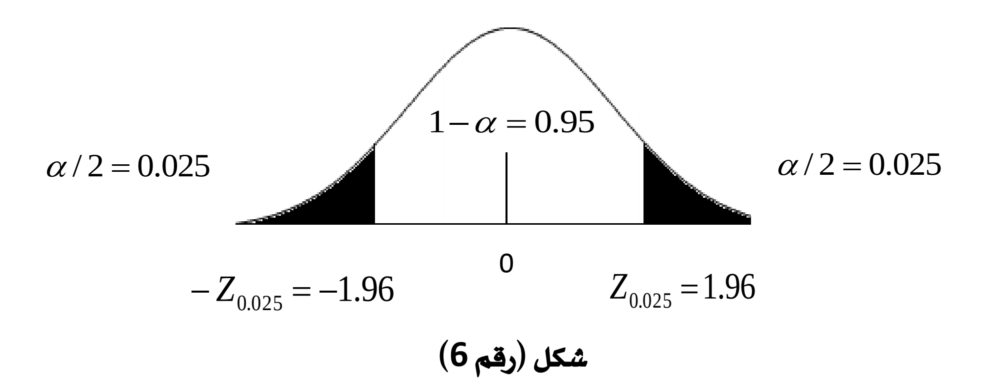
وأخيرا وبضرب حدود المعادلة في (1-) وتغيير المتساويات نحصل على المطلوب.

$$p(\overline{X}-Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}}) < \mu < \overline{X}+Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})) = 1-\alpha$$

- **مثال (17)**: كانت درجات الطلاب في إحدى الكليات تتوزع توزيعا قريبا من التوزيع الطبيعي. أخذت عينة عشوائية حجمها 81 طالبا من طلبة هذه الكلية فوجد أن الوسط الحسابي لدرجات الطلاب في العينة 70 درجة، والمطلوب تحديد فترة ثقة 95% ، 99% للوسط الحسابي لدرجات الطلاب في الكلية علما بان التباين لدرجات الطلاب في الكلية يساوي 100.

الحل:

(أ) عند مستوى الثقة 95% فان: $\alpha = 5\%$ ، وان: $\alpha/2 = 0.025$ ، والقيمة المعيارية (باستخدام الجداول بالملحق) هي: $Z_{\alpha/2} = Z_{0.025} = 1.96$ والشكل التالي يوضح ذلك:



شكل (رقم 6)

وعليه نجد أن حدود الثقة للمتوسط μ يمكن إيجادها من المعادلة التالية:

$$p\left[70-1.96\sqrt{\frac{100}{81}} < \mu < 70+1.96\sqrt{\frac{100}{81}}\right] = 0.95$$

$$p(67.82 < \mu < 72.18) = 0.95$$

وهذا يعني أن متوسط درجة الطلاب في هذه الكلية (أي قيمة μ) تقع في الحدين (67.82 , 72.18) بدرجة ثقة 95% وبنسبة خطأ مسموح به يساوي 5%.

(ب) عند مستوى الثقة 99% فان $\alpha = 1\%$، وان $\alpha / 2 = 0.005$، والقيمة المعيارية باستخدام الجداول هي: $Z_{\alpha/2} = Z_{0.005} = 2.58$ وبالتالي فان حدود الثقة تكون على الشكل التالي:

$$p\left[70-2.58\sqrt{\frac{100}{81}} < \mu < 70+2.58\sqrt{\frac{100}{81}}\right] = 0.99$$

$$p(67.13 < \mu < 72.87) = 0.99$$

وهذا يعني إن متوسط درجة الطالب في هذه الكلية لا تزيد عن 72.87 ولا تقل عن 67.13 بدرجة ثقة 99% وبنسبة خطأ مسموح به 1%.

(ب) إذا كان تباين المجتمع σ^2 غير معلوم

كثيرا ما يواجه الباحث عند دراسة أحد المجتمعات الإحصائية المعتدلة أو القريبة من الاعتدال أن التباين غير معلوم ويكون من الصعب إيجاد قيمة هذا التباين من المجتمع الأصلي نظرا لطبيعة مفردات هذا المجتمع أو لصعوبة ذلك عمليا. ولتقدير متوسط المجتمع μ من إحصائية العينة $\overline{X}$ في حالة عدم معرفتنا لهذه المعلمة فإننا نقوم بحساب الإحصائية S^2 من العينة كتقدير نقطة غير متحيز لتباين المجتمع ثم نتبع نفس الخطوات السابقة في تقدير فترة الثقة على أن نفرق بين حالتين، بين أن يكون حجم العينة كبيرا $n \geq 30$ وبين أن يكون حجم العينة صغيرا ($n<30$).

الحالة الأولى: التباين غير معلوم وحجم العينة كبير

في ضل الأحجام الكبيرة للعينات المختارة عشوائيا من المجتمعات الأصلية، يمكن للباحث استخدام خصائص توزيع القيم المعيارية Z في تقدير أي معلمة من معالم المجتمع الحقيقي سواء كان التباين معروف أو غير معروف. وبالتالي فإن فترة ثقة 100% $(1-\alpha)$ للوسط الحسابي تكون كالتالي:

$$p(\overline{X} - Z_{\alpha/2}(\frac{S}{\sqrt{n}}) < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$$

- **مثال (18)**: من مجتمع طلبة كلية مزايا سحبت عينة حجمها 49 طالبا لدراسة مستوى الذكاء لطلبة الكلية. ولهذا الغرض عقد اختبار ذكاء فكان متوسط ذكاء طلاب العينة هو 110 درجة والانحراف المعياري لها 11.5 درجة. المطلوب حساب فترة ثقة 95% لمتوسط ذكاء طلبة الكلية الذي سحبت منه العينة.

الحل:

$$p(\bar{X} - Z_{\alpha/2}(\frac{S}{\sqrt{n}}) < \mu < \bar{X} + Z_{\alpha/2}(\frac{S}{\sqrt{n}})) = 1-\alpha$$

$$p(110 - 1.96(\frac{11.5}{\sqrt{49}}) < \mu < 110 + 1.96(\frac{11.5}{\sqrt{49}})) = 0.95$$

$$p(106.78 < \mu < 113.22) = 0.95$$

وعلى ذلك فإن درجات الذكاء لمجموعة الطلبة في هذه الكلية تنحصر بين 106.78 درجة، 113.22 درجة بثقة قدرها 95% وبخطأ مقبول قدره 5%.

الحالة الثانية: التباين غير معلوم وحجم العينة صغير

إذا كان التباين غير معلوم وكان حجم العينة المعتمد عليها في التحليل الإحصائي صغيرا، فلا يكون بمقدورنا استعمال جداول التوزيع الطبيعي لإيجاد معامل الثقة Z، حيث أن الاعتماد على خصائص التوزيع المعتدل المعياري Z قد يوقعنا في خطأ إحصائي يؤثر على صحة التقدير، ولذلك فنحن نستخدم المتغير t حيث:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

والذي يخضع لتوزيع t بدرجات حرية $n-1$، وعليه فإننا نستخدم جداول t بدلا من جداول المنحنى المعتدل المعياري.

نظرية (10)

إذا كانت $X_1,\ldots,X_n$ عينة عشوائية صغيرة من $N(\mu,\sigma^2)$، حيث σ^2,μ مجهولتان، فإن فترة ثقة للوسط μ هي:

$$p(\bar{X} - t_{\alpha/2,n-1}(\frac{S}{\sqrt{n}}) < \mu < \bar{X} + t_{\alpha/2,n-1}(\frac{S}{\sqrt{n}})) = 1-\alpha$$

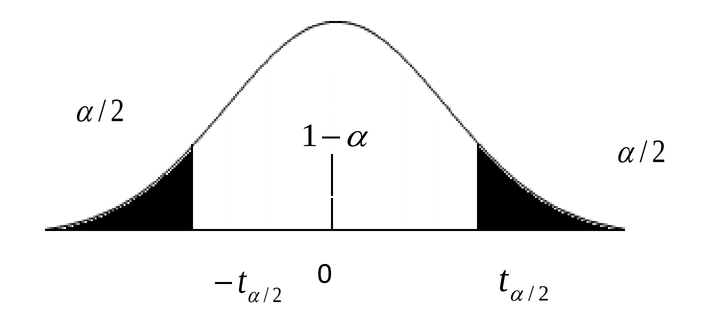
حيث: $\bar{X}$ = هو الوسط الحسابي للعينة.

S = الانحراف المعياري للعينة.

n = حجم العينة (العينة صغيرة).

$t_{\alpha/2,n-1} =$ هي قيمة t بدرجات حرية $n-1$ والتي تترك مساحة قدرها $\frac{\alpha}{2}$ على اليمين.

البرهان: من النظر إلى الشكل التالي:



شكل (رقم 7)

فإن احتمال أن تقع t بين القيمتين $t_{\alpha/2},\ t_{\alpha/2}$ - يساوي $1-\alpha$، أي أن:

$$p(-t_{\alpha/2} < t < t_{\alpha/2}) = 1-\alpha$$

وبالتعويض عن قيمة t بما يساويها في المعادلة أعلاه ينتج:

$$p(-t_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < t_{\alpha/2}) = 1-\alpha$$

وبضرب المعادلة بـ $\frac{S}{\sqrt{n}}$ ثم طرح $\overline{X}$ من كل حد، وأخيرا بالضرب في (1-) وعكس رموز المتباينة ينتج المطلوب.

- **مثال (19)**: باحتمال قدره 95 %، أوجد مدى الثقة لمتوسط المجتمع من بيانات العينة التالية:

3, 3, 5, 9, 10, 3, 4, 6, 9, 8

الحل:

حيث أن حجم العينة صغير ($n<30$) وأن التباين الخاص بالمجتمع غير معلوم فإننا نستخدم جداول t بعد أن نحسب القيمة المقدرة لتباين المجتمع من بيانات العينة كما يلي:

$$\bar{X} = \frac{\sum X}{n} = \frac{60}{10} = 6$$

$$S^2 = \frac{1}{n-1}(\sum X^2 - n\bar{X}^2)$$

$$= \frac{1}{10-1}(430 - 10(36)) = 7.78$$

وباستخدام جداول t نجد أن:

$$t_{\alpha/2,n-1} = t_{0.025,9} = 2.26$$

ومنها فإن:

$$p(6 - 2.26(\sqrt{\frac{7.78}{10}}) < \mu < 6 + 2.26(\sqrt{\frac{7.78}{10}})) = 95\%$$

وبذلك تكون حدود الثقة لمتوسط المجتمع في هذه الحالة هي: $4.006 < \mu < 7.994$

ثانيا: تقدير الفرق بين وسطين حسابيين لمجتمعين معتدلين مستقلين

نحتاج وفي حالات كثيرة إلى تقدير الفرق بين وسطين كتقدير الفرق بين ذكاء الطلاب والطالبات في إحدى الكليات أو تقدير الفرق بين أجور العمال والعاملات في أحد المصانع وذلك من خلال دراسة العينات، مع ملاحظة الحالات التالية:

(أ) إذا كانت تباينات المجتمع σ_2^2, σ_1^2 معلومة

نظرية (11)

إذا كانت $X_1,\ldots,X_n$ عينة عشوائية متوسطها $\bar{X}$ مختارة من توزيع طبيعي $N(\mu_1,\sigma_1^2)$، وكانت $Y_1,\ldots,Y_m$ عينة عشوائية ثانية متوسطها $\bar{Y}$ مختارة من

توزيع طبيعي $N(\mu_2, \sigma_2^2)$ مستقل عن التوزيع الأول، فإن فترة ثقة $(1-\alpha)100\%$ ثقة للفرق بين الوسطين $\mu_1 - \mu_2$ هي:

$$p\left[(\bar{X}-\bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} < \mu_1 - \mu_2 < (\bar{X}-\bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right] = 1-\alpha$$

البرهان:

$$\bar{X}-\bar{Y} \sim N(\mu_1 - \mu_2, \quad \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m})$$

لذا فإن:

$$Z = \frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$$

هي قيمة المتغير الطبيعي القياسي Z. وإن احتمال أن تقع Z بين القيمتين $-Z_{\alpha/2}, Z_{\alpha/2}$ هي:

$$p(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

فإذا عوضنا عن قيمة Z بما يساويها أعلاه ينتج:

$$p(-Z_{\alpha/2} < \frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} < Z_{\alpha/2}) = 1-\alpha$$

وبضرب كل حد بـ $\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}$ ثم بطرح $\bar{X}-\bar{Y}$ من كل حد، وأخيرا بضرب كل حد بـ(1-) ينتج المطلوب.

- **مثال (20): إذا كانت أعمار الطلاب والطالبات في إحدى الكليات تتوزع توزيعا معتدلا بانحراف معياري 5 ,7 على التوالي. ومن عينة تحتوي على 100 طالب وجد أن متوسط السن في العينة للطلاب 23 سنة. ومن عينة أخرى من الطالبات**

تحتوي على 144 طالبة وجد أن متوسط سن الطالبة 20 سنة. أوجد فترة ثقة للفرق بين المتوسطين في الكلية بدرجة ثقة 99%.

الحل:

من البيانات أعلاه نجد أن الحد الأدنى للفرق بين المتوسطين هو:

$$LL = (\bar{X} - \bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

$$= (23 - 20) - 2.58\sqrt{\frac{49}{100} + \frac{25}{144}} = 0.898$$

أما الحد الأعلى للفرق بين المتوسطين فهو:

$$UL = (\bar{X} - \bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

$$= (23 - 20) + 2.58\sqrt{\frac{49}{100} + \frac{25}{144}} = 5.10$$

أي أن الفرق بين أعمار الطلاب والطالبات في هذه الكلية يقع بين (0.898 , 5.10) سنة بدرجة ثقة 99% وبنسبة خطأ مقداره 1%.

(ب) إذا كانت تباينات المجتمع σ_2^2, σ_1^2 غير معلومة

في حالة كهذه فإننا نستخدم S_2^2, S_1^2 كتقديرين غير متحيزين لكل من σ_2^2, σ_1^2 على التوالي من بيانات كل من العينة الأولى والثانية، على أن نفرق مرة أخرى بين حجم العينات، أي بين أن يكون حجم العينات كبيرا ($n + m>50$) وبين أن يكون حجم العينات صغيرا ($n + m<50$) كما يلي:

الحالة الأولى: التباينات مجهولة وحجم العينات كبيرا

وهنا نستخدم S_2^2, S_1^2 كتقديرات غير متحيزة للتباينات المجهولة من بيانات العينتين بصرف النظر عن التوزيعات الأصلية طالما أن العينات كبيرة الأحجام ثم نستخدم جداول التوزيع الطبيعي المعياري Z.

نظرية (12)

إذا كانت $X_1,\dots,X_n$ عينة عشوائية كبيرة متوسطها $\bar{X}$ مأخوذة من توزيع طبيعي $N(\mu_1, \sigma_1^2)$ وكانت $Y_1,\dots,Y_m$ عينة عشوائية كبيرة أخرى مأخوذة من توزيع طبيعي $N(\mu_2, \sigma_2^2)$ مستقل عن التوزيع الأول، فإن فترة ثقة $(1-\alpha)100\%$ للفرق بين الوسطين $\mu_1 - \mu_2$ هي:

$$p\left[(\bar{X}-\bar{Y}) - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}} < \mu_1 - \mu_2 < (\bar{X}-\bar{Y}) + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}\right] = 1-\alpha$$

- **مثال (21):** سحبت عينة من مجتمع ما حجمها 36 مفردة فوجد أن وسطها الحسابي (118) وانحرافها المعياري 6. كما سحبت عينة أخرى من مجتمع ثاني مستقل عن المجتمع الأول حجمها 64 مفردة ووجد أن وسطها الحسابي 114 بانحراف معياري يساوي 7. أوجد فترة ثقة 95% للفرق بين الوسطين الحسابيين للمجتمعين.

الحل: من بيانات العينتين الآتية:

$$\bar{X} = 118, \quad S_1 = 6, \quad n = 36$$
$$\bar{Y} = 114, \quad S_2 = 7, \quad m = 64$$
$$1-\alpha = 0.95, \quad \alpha = 0.05, \quad \alpha/2 = 0.025$$

فإن فترة الثقة سوف تكون كالتالي:

$$p\left[(118-114) - 1.96\sqrt{\frac{36}{36}+\frac{49}{64}} < \mu_1 - \mu_2 < (118-114) + 1.96\sqrt{\frac{36}{36}+\frac{49}{64}}\right] = 0.95$$

$$p(1.4 < \mu_1 - \mu_2 < 6.6) = 95\%$$

وهذا يعني أن الفرق $\mu_1 - \mu_2$ لا يتعدى القيمة 6.6 ولا يقل عن القيمة 1.4 بدرجة ثقة 95% ونسبة خطأ $\alpha = 0.05$.

الحالة الثانية: التباينات مجهولة وحجم العينات صغير

نظرية (13)

إذا كان $\bar{X}$ هو الوسط الحسابي لعينة صغيرة حجمها n من مجتمع طبيعي وسطه μ_1 وتباينه σ_1^2، $\bar{Y}$ هو الوسط الحسابي لعينة صغيرة حجمها m مأخوذة من مجتمع طبيعي مستقل عن المجتمع الأول وسطه μ_2 وتباينه σ_2^2، وكانت التباينات المجهولة متساوية، أي أن $\sigma_1^2 = \sigma_2^2 = \sigma^2$، فإن فترة ثقة $(1-\alpha)100\%$ للفرق بين الوسطين $\mu_1 - \mu_2$ هي:

$$p\left[(\bar{X}-\bar{Y})-t_{\alpha/2,n+m-2}\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}<\mu_1-\mu_2<(\bar{X}-\bar{Y})+t_{\alpha/2,n+m-2}\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}\right]=1-\alpha$$

حيث S_p^2 هو التباين المجمع للعينتين معا ويعرف كالتالي:

$$S_p^2=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}$$

أما إذا كانت تباينات المجتمعين غير متساوية $(\sigma_1^2 \neq \sigma_2^2)$ فإن فترة ثقة $(1-\alpha)100\%$ للفرق بين الوسطين $\mu_1 - \mu_2$ هي:

$$p\left[(\bar{X}-\bar{Y})-t_{\alpha/2,d.f}\sqrt{(\frac{S_1^2}{n}+\frac{S_2^2}{m})}<\mu_1-\mu_2<(\bar{X}-\bar{Y})+t_{\alpha/2,d.f}\sqrt{(\frac{S_1^2}{n}+\frac{S_2^2}{m})}\right]=1-\alpha$$

حيث أن $d.f$ هي درجات الحرية (Degrees of freedom) وتساوي:

$$d.f=\frac{(\frac{S_1^2}{n}+\frac{S_2^2}{m})^2}{\frac{(S_1^2/n)^2}{n-1}+\frac{(S_2^2/m)^2}{m-1}}$$

مثال (22): عينة مكونة من 4 عمال اختيروا من مصنع معين فكانت أجورهم بالدولار (17 ,15 ,15 ,16). وعينة أخرى تحتوي على نفس العدد من العمال من مصنع آخر كانت أجورهم هي (17 ,16 ,14 ,15) ريالا. أفرض أن العينتين مستقلتان، والمجتمعين يتوزعان توزيعا قريبا من الاعتدال ولهما نفس التباين،، المطلوب:

(1) تقدير نقطة للفرق $\mu_1 - \mu_2$.

(2) تقدير الفرق بين متوسطي الأجور في المصنعين بصفة عامة عند $\alpha = 0.05$.

الحل: نقوم بحساب كل من $S_2^2, S_1^2, \bar{Y}, \bar{X}$ كتقديرات غير متحيزة للمعالم $\mu_1, \mu_2, \sigma_2^2, \sigma_1^2$

على التوالي، وسوف نجد أن:

$$\bar{X} = \frac{\sum X}{n} = \frac{17+15+15+16}{4} = 15.75$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{15+14+16+17}{4} = 15.5$$

$$S_1^2 = \frac{1}{n-1}(\sum x^2 - n\bar{x}^2) = \frac{1}{4-1}(995 - 4(15.75)^2) = 0.92$$

$$S_2^2 = \frac{1}{m-1}(\sum y^2 - n\bar{y}^2) = \frac{1}{4-1}(966 - 4(15.5)^2) = 1.67$$

وعليه فان:

(1) الفرق بين متوسطي العينتين $\bar{X}, \ \bar{Y}$ هو تقدير نقطة غير متحيز للفرق بين متوسطي المجتمعين $\mu_1 - \mu_2$ اللذين لهما نفس التباين، أي أن:

$$\bar{X} - \bar{Y} = 15.75 - 15.5 = 0.25$$

هو تقدير نقطة غير متحيز للمقدار $\mu_1 - \mu_2$.

(2) لإيجاد فترة الثقة فان علينا أن نحسب التباين المجمع أولا كما يلي:

$$S_p^2 = \frac{(4-1)(0.92)+(4-1)(1.67)}{4+4-2} = 1.295$$

وبذلك فان الحد الأدنى للثقة:

$$LL = (\bar{X} - \bar{Y}) - t_{\alpha/2,n+m-2}\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}$$

$$= (15.75 - 15.5) - t_{0.025,6}\sqrt{1.295(\frac{1}{4}+\frac{1}{4})}$$

$$= 0.25 - (2.447)(0.805) = -1.719$$

ويكون الحد الأعلى للثقة هو:

$$UL = (\bar{X} - \bar{Y}) + t_{\alpha/2,n+m-2}\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}$$

$$= (15.75 - 15.5) + t_{0.025,6}\sqrt{1.295(\frac{1}{4}+\frac{1}{4})}$$

$$= 0.25 + (2.447)(0.805) = 2.219$$

ونظرا لعدم وجود قيم بالسالب للأجور، فيكون الحد ألأدنى يساوي صفرا، وعلى ذلك فأن فترة الثقة تكون على الشكل التالي: $p(0 < \mu_1 - \mu_2 < 2.219) = 095$.

ثالثا: تقدير متوسط الفرق ($\mu_1 - \mu_2$) في حالة عدم استقلالية العينات

كان حديثنا حتى الآن عن تقدير فترة الثقة للفرق بين المتوسطات عندما تكون العينات مستقلة عن بعضها البعض. ونريد للحديث أن يدور الآن عن فترة الثقة للفرق بين المتوسطات عندما تكون العينات غير مستقلة. ويقال إن هذه العينة مستقلة عن تلك العينة إذا كانت مفردات إحداهما مستقلة تماما عن مفردات العينة الثانية، أما إذا كانت هناك علاقة بين العينتين فإن هذا يعني عدم الاستقلال كأن تكون المشاهدات في أزواج وهي بذلك تكون مرتبطة مع بعضها البعض. فإذا كان لدينا عينة من مجموعة مرضى ضغط الدم، ثم قمنا بقياس ضغط الدم لمفرداتها قبل وبعد تعاطي دواء معين، فأنه يكون لدينا عينتين ألأولى وهي القراءات قبل تعاطي الدواء والثانية هي

القراءات بعد تعاطي الدواء وهنا يكون عدد مفردات العينة الأولى متساوي مع عدد مفردات العينة الثانية وأن العينتين غير مستقلتين. أما إذا أخذنا مجموعة من الأشخاص وتم علاجهم بمعالجة معينة ثم أخذنا مجموعة أخرى من الأشخاص وتم علاجهم بمعالجة أخرى فإن العينتين في هذه الحالة مستقلتان وليس من الضرورة أن تكونا من نفس الحجم.

والآن إذا كان لدينا عينة حجمها n_1 غير مستقلة عن عينة أخرى حجمها n_2 فإن عدد مفردات العينتين غير مختلفين أي أن $n_1 = n_2 = n$، ثم يحسب الفرق بين المشاهدات المزدوجة حيث: $(d_i, \quad i = 1,\ldots,n)$

$d_i =$ **قيمة المفردة في العينة الثانية** (Y_i) $-$ **قيمة المفردة في العينة الأولى** (X_i)

فإذا كان $d_1,\ldots,d_n$ هو الفرق بين n من أزواج المشاهدات فإن تقدير النقطة لـ μ_d (في إشارة للفرق $\mu_1 - \mu_2$) هو $\bar{d}$، حيث $\bar{d}$ هو الوسط الحسابي للفروق والذي تعرفه المعادلة التالية:

$\bar{d} = \frac{\sum d_i}{n}$ وأن المتغير t حيث:

$$t = \frac{\bar{d}\sqrt{n}}{S_d}$$

يتبع توزيع t بدرجات حرية $n - 1$ حيث S_d هو الانحراف المعياري لهذه الفروق والذي يعرف كالتالي:

$$S_d = \sqrt{\frac{1}{n-1}\left(\sum d^2 - n\bar{d}^2\right)}$$

وبالتالي فإن فترة ثقة $(1-\alpha)100\%$ للفرق μ_d للأزواج من المشاهدات هي:

$$p\left(\bar{d} - t_{\alpha/2,n-1}\frac{S_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{\alpha/2,n-1}\frac{S_d}{\sqrt{n}}\right) = 1-\alpha$$

- **مثال (23)**: سحبت عينة من توزيع طبيعي $N(\mu,\sigma^2)$ تتكون من 5 أطفال رضع، سجلت أوزانهم فكانت بالكيلو غرام على النحو التالي:

12 ,13 ,9 ,10 ,11. ثم أعطيت مفردات هذه العينة غذاء من نوع جديد عدة أشهر وسجلت أوزانهم فكانت على النحو التالي: 12.5 ,13.2 ,10 ,10.3 ,11.9.

والمطلوب: تقدير متوسط الزيادة بصفة عامة نتيجة تناول الغذاء الجديد عند مستوى معنوية قدرها 1%.

الحل: الجدول التالي يبين خطوات الحل مع ملاحظة أن العينات غير مستقلة.

جدول (رقم 18)

الأطفال	الوزن قبل تناول الغذاء الجديد	الوزن بعد تناول الغذاء الجديد	الفرق في الوزن (d_i)	d_i^2
1	12	12.5	- 0.5	0.25
2	13	13.2	- 0.2	0.04
3	9	10	- 1.0	1.00
4	10	10.3	- 0.3	0.09
5	11	11.9	- 0.9	0.81
المجموع			**- 2.9**	**2.19**

ومن الجدول نجد أن:

$$\bar{d} = \frac{\sum d}{n} = \frac{-2.9}{5} = -0.58$$

$$S_d^2 = \frac{1}{n-1}(\sum d^2 - n\bar{d}^2)$$
$$= \frac{1}{5-1}(2.19 - 5(0.58))^2 = 0.127$$

$$S_d = \sqrt{0.127} = 0.356$$

وعليه فإن الحد الأدنى للثقة لـ μ_d هو:

$$LL = \bar{d} - t_{\alpha/2,n-1}\frac{S_d}{\sqrt{n}}$$

$$= -0.58 - 4.6(\frac{0.356}{\sqrt{5}}) = -1.31$$

أما الحد الأعلى للثقة فهو:

$$UL = \bar{d} + t_{\alpha/2,n-1}\frac{S_d}{\sqrt{n}}$$

$$= -0.58 + 4.6(\frac{0.356}{\sqrt{5}}) = 0.15$$

وهذا يعني أن:

$$p(-1.31 < \mu_d < 0.15) = 0.99$$

بمعنى أن فرق الزيادة في الوزن في المتوسط لن يتعدى 0.15 كغم ولن يقل عن -1.31 من الكيلوغرام، وذلك بحالة خطأ واحدة لكل مائة عينة. وبما أن هذه الفترة تحتوي على الصفر فمعنى ذلك أنه لا يوجد فرق بين الطريقتين في زيادة الوزن.

رابعا: تقدير نسبة المجتمع p

في حالات كثيرة نكون بصدد مجتمع ينقسم إلى قسمين، واحد يمتلك صفة معينة وآخر لا يمتلك تلك الصفة، كتقسيم مجتمع إنتاج أحد الآلات إلى معيب وآخر غير معيب. ونظرا لصعوبة تحديد نسبة الصفة التي تمتلكها مفردات المجتمع لأسباب ذكرناها سابقا، فإنه يمكن الاستفادة من خصائص توزيع النسب في العينات الممكنة المسحوبة من المجتمع وذلك في تقدير نسبة صفة ما في المجتمع من نسبتها في العينة. فإذا رمزنا للنسبة في المجتمع بالرمز p وللنسبة في العينة بالرمز $\hat{p}$ ، فسوف نكون بصدد تقدير المعلمة p من الإحصائية $\hat{p}$ وهو تقدير نقطة غير متحيز للمعلمة p. أما فترة ثقة $(1-\alpha)100\%$ للنسبة في المجتمع فهي:

$$p(\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = 1-\alpha$$

شرط أن تكون n كبيرة. أما إذا كانت n صغيرة فإن فترة الثقة للمعلمة p في المجتمع سوف تكون كما يلي:

$$p(\hat{p}-t_{\alpha/2,n-1}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}<p<\hat{p}+t_{\alpha/2,n-1}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}})=1-\alpha$$

- **مثال (24)**: أرادت إحدى الكليات معرفة رأي طلبتها في الأسلوب التدريسي لأحد أساتذتها، ولهذا الغرض أجرت استطلاعا بين عينة من طلابها حجمها 400 طالبا فوجدت أن عدد المؤيدين منهم لأسلوب الأستاذ الجامعي 320 طالبا. أوجد فترة ثقة 95% لنسبة الطلبة المؤيدين لأسلوب الأستاذ في الكلية.

الحل:

$$\hat{p}=\frac{320}{400}=0.8$$

ومنها فإن: $1-\hat{p}=0.20$، وأن $Z_{\alpha/2}=Z_{0.025}=1.96$، وبذلك يكون الحد الأدنى لنسبة المؤيدين في الكلية يساوي:

$$LL=\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$=0.80-1.96\sqrt{\frac{(0.8)(0.2)}{400}}=0.76$$

والحد الأعلى يساوي:

$$UL=0.8+1.96\sqrt{\frac{(0.8)(0.2)}{400}}=0.84$$

وعلى ذلك فإن نسبة المؤيدين لأسلوب الأستاذ التدريسي في الكلية يقع بين (0.76, 0.84) وذلك بدرجة ثقة 95% ونسبة خطأ 5%.

- **مثال (25)**: من دراسة لأحد الباحثين ثم سحب عينة من 20 طالبا في كلية الهندسة، فوجد أن عدد الذين لا يدخنون 5 طلاب فقط، والمطلوب:
  (أ) تقدير نقطة لنسبة المدخنين في الكلية بصفة عامة.
  (ب) تقدير فترة للمعلمة p عند $\alpha=0.05$.

الحل:

(أ) إن تقدير النقطة الغير متحيز للمعلمة p هو $\hat{p}$ وهي نسبة المدخنين في العينة، وعلى ذلك فإن: $\hat{p}=\frac{15}{20}=0.75$.

(ب) عند تقدير فترة للمعلمة p نجد أن حجم العينة صغير وأن $\alpha=5\%$ وبالتالي فإن: $t_{\alpha/2,n-1}=t_{0.025,19}=2.09$

وعليه فإن الحد الأدنى يساوي:

$$LL=\hat{p}-t_{\alpha/2,n-1}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$$
$$=0.75-2.09\sqrt{\frac{(0.75)(0.25)}{19}}=0.54$$

أما الحد ألأعلى

فهو:

$$UL=\hat{p}+t_{\alpha/2,n-1}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$$
$$=0.75+2.09\sqrt{\frac{(0.75)(0.25)}{19}}=0.96$$

وهذا يعني أن نسبة المدخنين في الكلية بشكل عام تقع بين 54% ، 96% بدرجة ثقة 95% .

خامسا: تقدير الفرق بين نسبتين في مجتمعين (p_1-p_2)

من المسائل التي تثير اهتمام الباحث الاجتماعي هي معرفة الفرق بين ظاهرة ما في أحد المجتمعات ونسبتها في مجتمع آخر، كمعرفة الفرق بين نسبة الأمية في مجتمع الذكور ومجتمع الإناث في سن معينة أو الفرق بين نسبة تغيب الطلاب والطالبات في أحد المواد. وفي حالات كهذه فإنه يكون من العملي سحب عينة من المجتمع الأول وتحديد النسبة فيها وعينة من المجتمع الآخر المستقل عن المجتمع الأول وتحديد

النسبة فيها أيضا لنقوم أخيرا بتقدير فترة ثقة للفرق بين النسبتين في المجتمعين $p_1 - p_2$ من الفرق بين النسبتين في العينتين $\hat{p}_1 - \hat{p}_2$ باستخدام فترة الثقة التالية:

$$p\left[(\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}\right] = 1-\alpha$$

حيث أن $\hat{p}_1$ هي نسبة الظاهرة في العينة الأولى التي حجمها n ، $\hat{p}_2$ هي نسبة الظاهرة في العينة الثانية التي حجمها m، وأن العينات كبيرة و $Z_{\alpha/2}$ قيمة المتغير الطبيعي القياسي الذي يترك مساحة $\alpha/2$ إلى اليمين.

البرهان:

نحن نعلم أن توزيع المعاينة لـ $\hat{p}_1 - \hat{p}_2$ عندما تكون n كبيرة يكون قريبا من التوزيع الطبيعي بوسط حسابي قدره $p_1 - p_2$ وانحراف معياري قدره $\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}$ ، لذا فإن:

$$p = (-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

حيث أن:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}}$$

وبالتعويض عن قيمة Z يكون:

$$p(-Z_{\alpha/2} < \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}} < Z_{\alpha/2}) = 1-\alpha$$

وبضرب كل حد بـ $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}$ وطرح $(\hat{p}_1 - \hat{p}_2)$ من كل حد ثم بضرب كل حد بـ (-1) ينتج المطلوب.

- **مثال (26)**: من عينة تحتوي على 100 طالب من طلبة جامعة النهرين وجد أن منهم 65 طالبا مدخنا. ومن عينة أخرى من طلبة جامعة بغداد بنفس الحجم وجد أن منهم 45 طالبا مدخنا. أوجد فترة ثقة 99% للفرق بين نسبتي الطلبة المدخنين في الجامعتين.

الحل $\hat{p}_1$ = 65% وهي نسبة الطلبة المدخنين في جامعة النهرين.

$\hat{p}_2$ = 45% وهي نسبة الطلبة المدخنين في جامعة بغداد.

$$1-\alpha = 99\%$$

$$Z_{\alpha/2} = Z_{0.005} = 2.58$$

وبالتالي فإن الحد الأدنى للفرق بين النسبتين في الجامعتين هو:

$$LL = (\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$

$$= (0.65 - 0.45) - 2.58\sqrt{\frac{(0.65)(0.35)}{100} + \frac{(0.45)(0.55)}{100}} = 0.022$$

أما الحد الأعلى للفرق بين النسبتين في الجامعتين فهو:

$$UL = (0.65 - 0.45) + 2.58\sqrt{\frac{(0.65)(0.35)}{100} + \frac{(0.45)(0.55)}{100}} = 0.378$$

وهذا يعني أن الفرق بين النسبتين في الجامعتين يقع بين 0.022 كحد أدنى و 0.378 كحد أقصى بدرجة ثقة 99% ونسبة خطأ α تساوي 1%.

سادسا: تقدير تباين المجتمع σ^2

إذا كان لدينا عينة حجمها n وتباينها S^2 مسحوبة عشوائيا من مجتمع يتوزع توزيعا معتدلا أو قريبا من الاعتدال بوسط μ وتباين σ^2 فإن الإحصائية $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ تخضع لتوزيع χ^2 بدرجات حرية n-1 وهذا يعني أن:

$$p(\chi^2_{1-\alpha/2,n-1} \le \chi^2 \le \chi^2_{\alpha/2,n-1}) = 1-\alpha$$

وبالتعويض عن قيمة χ^2 بما يساويها في أعلاه ينتج:

$$p(\chi^2_{1-\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}) = 1-\alpha$$

ومنها فإن فترة الثقة لتباين معلمة σ^2 سوف تكون على الشكل التالي:

$$p(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}) = 1-\alpha$$

أما إذا أردنا تقدير فترة ثقة للمعلمة σ فإنها تكون على الشكل التالي:

$$p(\frac{(\sqrt{n-1})S}{\sqrt{\chi^2_{\alpha/2,n-1}}} \le \sigma \le \frac{(\sqrt{n-1})S}{\sqrt{\chi^2_{1-\alpha/2,n-1}}}) = 1-\alpha$$

- **مثال (27)**: أخذت عينة من 10 طلاب فوجد أن الوسط الحسابي لأطوالهم يساوي 170 سم والانحراف المعياري 2.1. المطلوب تحديد فترة ثقة لمعلمة المجتمع σ^2 للأطوال في المجتمع الطلابي كله عند $\alpha = 5\%$.

الحل: باستخدام جداول χ^2 نجد أن:

$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,9} = 19.02$$
$$\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,9} = 2.70$$

وعليه فإن الحد الأدنى للثقة:

$$LL = \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} = \frac{9(4.4)}{19.02} = 2.09$$

أما الحد الأعلى للثقة فهو:

$$UL = \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} = \frac{9(4.4)}{2.70} = 14.7$$

سابعا: تقدير فترة ثقة للنسبة $\frac{\sigma_1^2}{\sigma_2^2}$

أفرض أن $X_1,\ldots,X_n$ عينة عشوائية تباينها S_1^2 مسحوبة من مجتمع $N(\mu_1,\sigma_1^2)$، وأن $Y_1,\ldots,Y_m$ عينة عشوائية ثانية تباينها S_2^2 مستقلة عن الأولى ومسحوبة أيضا من مجتمع $N(\mu_2,\sigma_2^2)$ فإن المتغير

$$F=\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}=\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$$

يخضع لتوزيع *F* على درجات الحرية *(n-1),(m-1)*، وعليها فإن فترة ثقة $(1-\alpha)100\%$ للنسبة $\frac{\sigma_1^2}{\sigma_2^2}$ هي:

$$p(\frac{S_1^2}{S_2^2}.\frac{1}{f_{\alpha/2,(n-1,m-1)}}\leq\frac{\sigma_1^2}{\sigma_2^2}\leq\frac{S_1^2}{S_2^2}.\frac{1}{f_{1-\alpha/2,(n-1,m-1)}})=1-\alpha$$

حيث $\frac{S_1^2}{S_2^2}$ هو تقدير نقطة للمعلمة $\frac{\sigma_1^2}{\sigma_2^2}$. ولأن جداول *F* لا تحتوي على قيم *F* عند درجات الحرية المحددة و $1-\alpha$ لذلك فانه يمكن استخدام العلاقة التالية:

$$f_{1-\alpha/2,(n-1,m-1)}=\frac{1}{f_{\alpha/2,(m-1,n-1)}}$$

- **مثال (28):** سحبت عينة حجمها 6 من توزيع $N(\mu_1,\sigma_1^2)$، فكان تباينها يساوي 144. كما سحبت عينة أخرى مستقلة عن العينة الأولى حجمها 10 من توزيع $N(\mu_2,\sigma_2^2)$ فكان تباينها 49. قدر فترة ثقة 90٪ للنسبة $\frac{\sigma_1^2}{\sigma_2^2}$ ؟

الحل: باستخدام جداول F، فإن الحد الأدنى للثقة هو:

$$LL = \frac{S_1^2}{S_2^2} \cdot \frac{1}{f_{\alpha/2,(n-1,m-1)}}$$
$$= \frac{144}{49}\left(\frac{1}{3.48}\right) = 0.84$$

أما الحد الأعلى فهو:

$$UL = \frac{S_1^2}{S_2^2} \cdot f_{\alpha/2(m-1,n-1)}$$
$$= \frac{144}{49}(4.87) = 14.32$$

وهذا يعني أن $14.32 > \frac{\sigma_1^2}{\sigma_2^2} > 0.87$ بدرجة ثقة 90%.

ثامنا: فترات الثقة لمعالم الانحدار α, β

من المعروف ان العلاقة الدالية بين المتغيرين X, Y يمكن أن تكتب على شكل نموذج خطي يأخذ الصيغة التالية:

$$Y_i = \alpha + \beta X_i + e_i , \qquad i = 1,2,\ldots,n$$

وأن الإحصائيتين $\hat{\alpha}, \hat{\beta}$ هما تقديرا نقطة غير متحيزين للمعالم α, β في خط الانحدار المستقيم. ولكي نتمكن من إيجاد فترات الثقة المطلوبة فإننا نفترض أن الأخطاء العشوائية مستقلة عن بعضها البعض وتخضع للتوزيع الطبيعي. فإذا كانت العينة من الأزواج المرتبة $[(X_1,Y_1),\ldots,(X_n,Y_n)]$ والمسحوبة من المجتمع تتوزع توزيعا معتدلا فإن الإحصائية:

$$t = \frac{\hat{\beta} - \beta}{S(\hat{\beta})}$$

تخضع لتوزيع t بدرجات حرية $n-2$، حيث $S(\hat{\beta})$ هو الانحراف المعياري للإحصائية $\hat{\beta}$. وهذا يعني أنه الجذر التربيعي لتباين الإحصائية $\hat{\beta}$ ، حيث:

$$S^2(\hat{\beta}) = \frac{S^2_{y/x}}{\sum(X-\bar{X})^2} = \frac{S^2_{y/x}}{\sum X^2 - n\bar{X}^2}$$

علما بأن $S^2_{y/x}$ هو الخطأ المعياري للتقدير والذي يعرف كما يلي:

$$S^2_{y/x} = \frac{\sum(y-\hat{y})^2}{n-2}$$

ومنها فإن $S(\hat{\beta}) = \sqrt{S^2(\hat{\beta})}$.

كما أن

$$t = \frac{\hat{\alpha}-\alpha}{S(\hat{\alpha})}$$

تمثل هي الأخرى متغيرا عشوائيا له التوزيع t بدرجات الحرية $n-2$، علما بأن $S(\hat{\alpha})$ هي الانحراف المعياري للإحصائية $\hat{\alpha}$، أي أنه الجذر التربيعي لتباين الإحصائية $\hat{\alpha}$، حيث:

$$S^2_{(\hat{\alpha})} = \frac{S^2_{Y/X}(\sum X^2)}{n(\sum X^2 - n\bar{X}^2)}$$

ومنها نجد أن $S(\hat{\alpha}) = \sqrt{S^2(\hat{\alpha})}$. وباستعمال هذه البيانات نجد أن فترة ثقة $(1-\alpha)100\%$ للمعلمة α هي:

$$p(-t_{\alpha/2,n-2} \le \frac{\hat{\alpha}-\alpha}{S(\hat{\alpha})} \le t_{\alpha/2,n-2}) = 1-\alpha$$

وبضرب جميع الأطراف في $S(\hat{\alpha})$ ثم طرح $\hat{\alpha}$ من جميع الأطراف وأخيرا بضرب الأطراف في (1-) مع تغيير المتراجحات نجد أن:

$$p(\hat{\alpha} - t_{\alpha/2,n-2}\ S(\hat{\alpha}) \le \alpha \le \hat{\alpha} + t_{\alpha/2,n-2} S(\hat{\alpha})) = 1-\alpha$$

وبذات الأسلوب نجد أن فترة ثقة $(1-\alpha)100\%$ للمعلمة β والتي تمثل ميل خط الانحدار كما يلي:

$$p(-t_{\alpha/2,n-2} \leq \frac{\hat{\beta}-\beta}{S(\hat{\beta})} \leq t_{\alpha/2,n-2}) = 1-\alpha$$

وبإجراء العمليات الجبرية السابقة نحصل على:

$$p(\hat{\beta}-t_{\alpha/2,n-2}\ S(\hat{\beta}) \leq \beta \leq \hat{\beta}+t_{\alpha/2,n-2}S(\hat{\beta})) = 1-\alpha$$

مثال (29): يمثل الجدول الآتي أطوال 10 أشجار من نوع معين (Y)، وعمر كل شجرة منها (X).

جدول (رقم 19)

الطول (قدم)	33	21	29	35	25	20	23	28	34	21
العمر (سنة)	13	9	13	12	10	9	10	17	12	8

والمطلوب:

(1) أوجد تقديرات النقطة لكل من β,α في معادلة خط انحدار الطول على العمر.

(2) أوجد فترات ثقة 95% للمعلمتين α,β في خط الانحدار المقدر.

الحل: (1) الجدول التالي يمثل خطوات حساب معادلة خط انحدار Y على X.

جدول (رقم 20)

Y	X	YX	X^2
33	13	429	169
21	9	189	81
29	13	377	169
35	12	420	144
25	10	250	100
20	9	180	81
23	10	230	100
38	17	646	289
34	12	408	144
21	8	168	64
279	**113**	**3297**	**1341**

$$\overline{X}=\frac{\sum X}{n}=\frac{113}{10}=11.3,\quad \overline{Y}=\frac{\sum Y}{n}=\frac{279}{10}=27.9$$

$$\hat{\beta}=\frac{\sum XY-n\overline{X}\overline{Y}}{\sum X^2-n\overline{X}^2}=\frac{3297-10(11.3)(27.9)}{1341-10(11.3)^2}=2.25$$

وهو تقدير لمعامل الانحدار β .

$$\hat{\alpha}=\overline{Y}-\hat{\beta}\overline{X}=27.9-(2.25)(11.3)=2.475$$

وهو تقدير نقطة للجزء المقطوع α .

(2) لإنشاء فترة ثقة يجب عمل تقدير للتباين المجهول $\sigma^2_{Y/X}$ ، ولكننا نعلم أن $S^2_{Y/X}$ هو تقدير غير متحيز للتباين $\sigma^2_{Y/X}$ حيث أن:

$$S^2_{Y/X}=\frac{\sum(Y-\hat{Y})^2}{n-2}$$

والذي يمكن كتابته بالشكل البسيط الآتي:

$$S_{Y/X}^2 = \frac{S_{yy} - \hat{\beta}^2 S_{xx}}{n-2}$$

علما بأن:

$$S_{yy} = \sum Y^2 - n\bar{Y}^2$$
$$= 8191 - 10(27.9)^2 = 406.9$$
$$S_{xx} = \sum X^2 - n\bar{X}^2$$
$$= 1341 - 10(11.3)^2 = 64.1$$
$$\therefore S_{y/x}^2 = \frac{406.9 - (2.25)^2(64.1)}{10-2} = 10.299$$

وبما أن فترة الثقة للمعلمة α هي:

$$p(\hat{\alpha} - t_{\alpha/2,n-2}S(\hat{\alpha}) \le \alpha \le \hat{\alpha} + t_{\alpha/2,n-2}S(\hat{\alpha})) = 1-\alpha$$

فانه وباستخدام جداول t نجد أن: $t_{\alpha/2,n-2} = t_{0.025,8} = 2.31$، كما أن:

$$S(\hat{\alpha}) = \sqrt{\frac{S_{y/x}^2(\sum X^2)}{n(\sum X^2 - n\bar{X}^2)}} = \sqrt{\frac{(10.299)(1341)}{10(64.1)}} = 4.64$$

وبالتالي

$$p(2.475 - (2.31)(4.64) \le \alpha \le 2.475 + (2.31)(4.64)) = 0.95$$

$$p(-8.24 \le \alpha \le 13.19) = 0.95$$

وهذه هي فترة الثقة للمعلمة α باحتمال قدره 95%.

أما فترة الثقة للمعلمة β فهي:

$$p(\hat{\beta} - t_{\alpha/2,n-2}S(\hat{\beta}) \le \beta \le \hat{\beta} + t_{\alpha/2,n-2}S(\hat{\beta})) = 1-\alpha$$

حيث

$$S(\hat{\beta}) = \sqrt{\frac{S^2_{y/x}}{\sum X^2 - n\bar{X}^2}} = \sqrt{\frac{(10.299)}{64.1}} = 0.4$$

إذن

$$p(2.25-(2.31)(0.4) \le \beta \le 2.25+(2.31)(0.4)) = 0.95$$

$$p(1.326 \le \beta \le 3.174) = 0.95$$

وهي فترة الثقة المطلوبة باحتمال 95%.

تاسعا: تقدير فترة ثقة لمعامل الارتباط ρ

إن معامل الارتباط r لعينة عشوائية حجمها n هو تقدير نقطة غير متحيز لمعامل الارتباط في المجتمع ρ ، والذي يحسب بالمعادلة التالية:

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2}\sqrt{\sum y^2 - n\bar{y}^2}}$$

ولإيجاد تقدير فترة ثقة لمعامل ارتباط المجتمع ρ (شرط أن $\rho \neq 0$) فان توزيع المعاينة لمعامل ارتباط العينة r معقد جدا ولا يمكن استخدامه لإيجاد فترة الثقة.

ولتحقيق هذا الهدف فقد قام *Fisher* بتحويل قيم r إلى قيم Z بالمعادلة التالية:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

ووجد أن Z تقترب من التوزيع الطبيعي بوسط حسابي قدره $\mu_z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}$

وتباين قدره $\sigma^2_z = 1/n-3$ وهذا معناه أن المتغير z :

$$z = \frac{Z-\mu_z}{\sigma_z}$$

يخضع لتوزيع طبيعي معياري $N(0,1)$ شرط أن يكون حجم العينة كبيرا.

وقد تم إعداد جداول خاصة بهذا التحويل (أنظر الملحق)لاستخدامه مباشرة بدلا من استخدام المعادلة أعلاه. ولكي نجد فترة ثقة $(1-\alpha)100\%$ لمعامل الارتباط في

المجتمع ρ فإننا نقوم بإيجاد فترة ثقة لـ μ_z ثم تحويلها إلى فترة ثقة للمعلمة ρ كما يلي:

$$p(-Z_{\alpha/2} \leq \frac{Z-\mu_z}{\sigma_z} \leq Z_{\alpha/2}) = 1-\alpha$$

ومنه نستنتج أن:

$$p(Z - Z_{\alpha/2}\frac{1}{\sqrt{n-3}} \leq \mu_z \leq Z + Z_{\alpha/2}\frac{1}{\sqrt{n-3}}) = 1-\alpha$$

ومن جداول (رقم 6) نحول حدي فترة الثقة أعلاه لـ μ_z إلى فترة ثقة للمعلمة ρ .

إما إذا كانت العينات صغيرة فإننا نضع $t_{\alpha/2,n-3}$ بدلا من $Z_{\alpha/2}$.

- **مثال (30)**: احسب تقدير نقطة لمعامل الارتباط ρ للبيانات الواردة في مثال (29) ثم باحتمال قدره 95% احسب تقدير فترة ثقة لـ ρ .

الحل:

لإيجاد تقدير نقطة لمعامل الارتباط ρ فإننا نقوم بحساب معامل الارتباط من العينة r كما يلي:

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2}\sqrt{\sum y^2 - n\bar{y}^2}}$$

$$r = \frac{3297 - 10(11.3)(27.9)}{\sqrt{1341 - 10(11.3)^2}\sqrt{8191 - 10(27.9)^2}} = 0.89$$

إما فترة ثقة 95% لـ ρ فانه عندما تكون $r = 0.89$ فان قيمة Z المقابلة لها هي:

$$Z = \frac{1}{2}\ln\frac{1+0.89}{1-0.89} = 1.422$$

هذه القيمة يمكن استخراجها من جداول تحويل r إلى Z كما أن:

$\sigma_z = \frac{1}{\sqrt{10-3}} = 0.378$، ولأن حجم العينة صغير فان: $t_{\alpha/2,n-3} = t_{0.025,7} = 2.37$.

وبتعويض هذه القيم في فترة الثقة نجد أن:

$$p(Z - t_{\alpha/2,n-3}(\sigma_z) \le \mu_z \le Z + t_{\alpha/2,n-3}(\sigma_z)) = 1-\alpha$$

$$p(1.422 - (2.37)(0.378) \le \mu_z \le 1.422 + (2.37)(0.378)) = 0.95$$

$$p(0.53 \le \mu_z \le 2.32) = 0.95$$

ومن جداول (رقم 6) نجد أن قيم ρ المناظرة لقيم μ_z هي:

$$0.4854 < \rho < 0.9809$$

وهي فترة ثقة تقريبية لـ ρ باحتمال 95%.

(4.2) خطأ المعاينة والدقة *Sampling Error and Precision*

في دراستنا السابقة تعرفنا على كيفية تقدير معالم المجتمع، غير أننا لم نتعرف على خطأ التقدير أو كمية ألأخطاء التي وقعنا فيها نتيجة تقدير معلمة المجتمع المجهولة. ويعرف خطأ المعاينة على انه الفرق بين متوسط العينة ومتوسط المجتمع إذا كنا بصدد تقدير متوسط المجتمع، فإذا كان متوسط المجتمع يساوي 7 وكان متوسط العينة المحسوبة يساوي 5 فان الفرق بينهما والذي يساوي 2 يعرف بخطأ المعاينة.

قاعدة عامة: إذا كانت معلمة المجتمع المجهولة هي θ ثم سحبنا عينة عشوائية من ذلك المجتمع وقدرنا قيمة المعلمة ورمزنا للتقدير بالرمز $\hat{\theta}$، فان خطأ المعاينة (Sampling Error) يحسب كالتالي:

$$SE = \hat{\theta} - \theta$$

وبموجب هذه المعادلة فقد يكون الخطأ موجبا أو سالبا وهو يتغير من عينة إلى أخرى. ولكن كيف نستطيع أن نحصل على خطأ المعاينة وان معلمة المجتمع θ غالبا ما تكون مجهولة ؟

وللإجابة على هذا السؤال علينا أن نعرف الدقة (*Precision*). وتعرف الدقة بأنها الحد الأقصى للفرق بين المعلمة والتقديرات المحسوبة من كل العينات الممكنة المسحوبة من مجتمع معين سواء بإرجاع أو بدون إرجاع.

ففي حالة تقدير المتوسط مثلا وجدنا أن فترة 95% ثقة هي:

$$p(\bar{X} - 1.96(\frac{\sigma}{\sqrt{n}}) \leq \mu \leq \bar{X} + 1.96(\frac{\sigma}{\sqrt{n}})) = 0.95$$

أي أن:

$$p(\left|\bar{X} - \mu\right| \leq 1.96(\frac{\sigma}{\sqrt{n}})) = 0.95$$

وبذلك تكون أقصى قيمة للفرق المطلق بين المعلمة والتقدير هي ($1.96(\frac{\sigma}{\sqrt{n}})$)، فإذا رمزنا لهذا الفرق (وهو الذي يعني الدقة) بالرمز D فانه يمكن أن نعرف الدقة بما يلي:

$$D = \left|\bar{X} - \mu\right| = 1.96(\frac{\sigma}{\sqrt{n}})$$

ومن هنا يتضح الفرق بين الدقة وخطأ المعاينة حيث أن خطأ المعاينة يحسب لعينة واحدة أما الدقة فهي الحد الأقصى لخطأ المعاينة لكل العينات المسحوبة من المجتمع.

قاعدة عامة: إذا كانت لدينا معلمة مجهولة وهي θ من مجتمع حجمه N ثم سحبت كل العينات من الحجم n سواء بإرجاع أو بدون إرجاع، فان دقة التقدير تحسب كالتالي:

$$D = \left|\hat{\theta} - \theta\right| = Z_{\alpha/2}(\sigma_{\hat{\theta}})$$

حيث $\sigma_{\hat{\theta}}$ هو الخطأ الخاص بالتقدير وان $Z_{\alpha/2}$ هي القيمة المعيارية المستخرجة من جداول المنحنى الطبيعي والتي يكون الاحتمال بعدها يساوي $\alpha/2$.

- **مثال (31):** سحبت عينة حجمها 81 مفردة من مجتمع حجمه 1000 مفردة وتباينه 144. أوجد دقة التقدير عندما يكون السحب بإرجاع وبدون إرجاع باحتمال قدره 99%.

الحل:

(أ) عندما يكون السحب بالإرجاع:

$$D = \left|\bar{X} - \mu\right| = Z_{\alpha/2}\,(\frac{\sigma}{\sqrt{n}})$$

$$= Z_{0.025}\sqrt{\frac{144}{81}} = (2.58)(\frac{12}{9}) = 3.44$$

(ب) عندما يكون السحب بدون إرجاع:

$$D = \left|\bar{X} - \mu\right| = Z_{\alpha/2}\,(\frac{\sigma}{\sqrt{n}})\sqrt{\frac{N-n}{N-1}}$$

$$= (2.58)(\frac{12}{9})\sqrt{\frac{1000-81}{1000-1}} = 3.30$$

ومن هنا يتضح أن كمية الخطأ في حالة السحب بإرجاع أكبر من كميته في حالة السحب بدون إرجاع. وفي حالة أن تكون σ غير معلومة فأنه يمكن إيجاد قيمة الدقة عن طريق حساب S من

أحد العينات واستخدامها كتقدير جيد للمعلمة المجهولة σ. وهنا تعرف الدقة كما يلي:

$$D = Z_{\alpha/2}\,(\frac{S}{\sqrt{n}})$$

(5.2) تقدير حجم العينة *Sample Size Estimation*

تعرفنا فيم سبق أن هناك علاقة بين الدقة وحجم العينة وهذا أمر طبيعي، فالدقة مثلا في حالة الوسط الحسابي كلما كانت صغيرة كلما تطلب ذلك حجما كبيرا للعينة والعكس صحيح. **ولكي نتعرف على هذه العلاقة سوف نقوم بتقدير حجم العينة في كل حالة من الحالات التالية:**

أولا: حالة السحب بالإرجاع:

في هذه الحالة، عرفنا الدقة عند تقدير متوسط المجتمع كما يلي:

$$D = \left|\overline{X} - \mu\right| = Z_{\alpha/2}\,(\frac{\sigma}{\sqrt{n}})$$

وبتربيع الطرفين نحصل على:

$$D^2 = (Z_{\alpha/2})^2\,(\frac{\sigma^2}{n})$$

ومنها نجد أن:

$$n = (\frac{Z_{\alpha/2}\,\sigma}{D})^2 \qquad (1)$$

حيث n تمثل حجم العينة في حالة السحب بالإرجاع عند تقدير المتوسط. **ونلاحظ من المعادلة (1) ما يأتي:**

أ) كلما كانت الدقة D صغيرة جدا فان معنى ذلك أن متوسط المقدر يساوي تقريبا متوسط المجتمع، وبالتالي فان حجم العينة n سوف يكون كبيرا جدا.

ب) كلما كبر الانحراف المعياري لمفردات المجتمع σ، كلما أدى ذلك إلى زيادة حجم العينة حتى تكون ممثلة للمجتمع أحسن تمثيل ذلك لأن كبر σ يعني أن المجتمع غير متجانس.

ج) كلما زاد الاحتمال كلما زادت تبعا لذلك قيمة $Z_{\alpha/2}$ وهذا يؤدي بدوره أيضا إلى زيادة حجم العينة.

- **مثال (32)**: في مجتمع يتكون من 500 وحدة، أوجد حجم العينة اللازم لتقدير الوسط الحسابي وذلك باحتمال % 95 وبدقة تساوي 0.9 ، إذا كان الانحراف المعياري للمجتمع يساوي 7 وكان السحب بإرجاع.

الحل:

$$n = (\frac{Z_{\alpha/2}\,\sigma}{D})^2 = (\frac{(Z_{0.025})(7)}{0.9})^2 = (\frac{(1.96)(7)}{0.9})^2 = 232$$

ثانيا: حالة السحب بدون إرجاع:

في هذه الحالة كنا قد عرفنا الدقة في حالة تقدير متوسط المجتمع كما يلي:

$$D = Z_{\alpha/2}\,(\frac{\sigma}{\sqrt{n_0}})\sqrt{\frac{N-n_0}{N-1}} = Z_{\alpha/2}\,\frac{\sigma}{\sqrt{n_0}}\sqrt{1-\frac{n_0}{N}}$$

حيث n_0 هو حجم العينة في حالة السحب بدون إرجاع. وبتربيع الطرفين وإيجاد قيمة n_0 نجد أن:

$$n_0 = \frac{N(Z_{\alpha/2}\,\sigma)^2}{ND^2 + (Z_{\alpha/2}\,\sigma)^2} \qquad (2)$$

- **مثال (33)**: في مثال (32)، احسب حجم العينة إذا كان السحب بدون إرجاع ؟

الحل: بتطبيق المعادلة (2) نجد أن:

$$n_0 = \frac{500((1.96)(7))^2}{500(0.9)^2 + ((1.96)(7))^2} = 159$$

من الملاحظ أن هناك علاقة بين حجم العينة في حالة السحب بإرجاع وبدون إرجاع، ويمكن إيجادها بقسمة الطرف الأيمن من المعادلة (2) على ND^2 لنحصل على:

$$n_0 = \frac{(Z_{\alpha/2}\,\sigma / D)^2}{1+(\dfrac{Z_{\alpha/2}\,\sigma}{D})/N} = \frac{n}{1+\dfrac{n}{N}} \qquad (3)$$

ومن هذه العلاقة يتضح أن حجم العينة في حالة ألإرجاع يكون أكبر من نظيره في حالة عدم الإرجاع. أما إذا كانت N كبيرة جدا فأن حجم العينة في كلا الحالتين يكون متساويا تقريبا.

- **مثال (34)**: بالعودة لمثال (32)، المطلوب دراسة تأثير زيادة حجم المجتمع على العينة في الحالات التالية:

(أ) إذا كان حجم المجتمع يساوي 10,000.

(ب) إذا كان حجم المجتمع يساوي 20,000.

(ج) إذا كان حجم المجتمع يساوي 100,000.

الحل:

(أ) إذا كان حجم المجتمع 10,000 فان حجم العينة سوف يكون كما يلي:

$$n_0 = \frac{232}{1+\dfrac{232}{10,000}} = 227$$

(ب) إذا كان حجم المجتمع 20,000 فان حجم العينة يساوي:

$$n_0 = \frac{232}{1+\dfrac{232}{20,000}} = 229$$

(ج) إذا كان حجم المجتمع 100,000 فان حجم المجتمع يساوي:

$$n_0 = \frac{232}{1+\dfrac{232}{100,000}} = 231$$

وهكذا يتضح أن حجم العينة في حالة السحب بدون إرجاع يقترب من حجم العينة في حالة السحب بالإرجاع حتى يتساويا كلما زاد حجم المجتمع شيئا فشيئا.

تمـــارين

(1) من مجتمع يتكون من المفردات التالية: 7 ,5 ,4 ,3 ,2 سحبت عينات من الحجم 2، والمطلوب:

(أ) تقدير متوسط المجتمع بنقطة ثم إثبات أن هذا التقدير تقدير غير متحيز.

(ب) قارن بين التقدير في حالة السحب بإرجاع والسحب بدون إرجاع واثبت أيهما أكفأ.

(ج) أوجد تقدير غير متحيز لمعلمة المجتمع σ^2 .

(د) أوجد حدود الثقة لمتوسط المجتمع باحتمال 95% .

(هـ) احسب خطأ المعاينة للعينات المختلفة ثم احسب دقة التقدير عندما يكون السحب بإرجاع وبدون إرجاع.

(2) في السؤال أعلاه:

(1) أوجد تقديرا لنسبة الأرقام الزوجية في المجتمع في حالة السحب بإرجاع وبدون إرجاع.

(2) قارن بين التقديرين من حيث عدم التحيز والكفاءة.

(3) أوجد حدود الثقة المختلفة لنسبة الأرقام الزوجية في المجتمع باحتمال 95% .

(3) من البيانات التالية والخاصة بعينة مسحوبة من مجتمع بإرجاع، المطلوب إيجاد حدود الثقة لمتوسط المجتمع باحتمال 99% : 11 ,9 ,10 ,6 ,8 ,7 ,6 ,4 ,2 ,1

(4) إذا كان الانحراف المعياري لمجتمع يتكون من 800 وحدة يساوي 7. أوجد حجم العينة اللازم لتقدير الوسط الحسابي وذلك باحتمال 99% وبدقة 0.01 في حالة السحب بإرجاع وبدون إرجاع.

(5) في مجتمع يتكون من 1000 وحدة، أوجد حجم العينة اللازم لتقدير نسبة المدخنين في المجتمع وذلك باحتمال 95% وبدقة مقدارها 3% علما بأن نسبة المدخنين في المجتمع هي 7%.

(6) لمقارنة تكاليف المعيشة لعائلتين من مدينتين مختلفتين تتكون كل عائلة منهما من 5 أفراد، أخذت عينتان الأولى حجمها 14 فردا بمتوسط قدره 26 دولارا من مجتمع $N(\mu, 2.25)$ ، والثانية حجمها 11 بمتوسط قدره 21 دولارا من مجتمع طبيعي آخر $N(\mu, 2.69)$.

والمطلوب 95% حساب فترة ثقة للفرق بين الوسطين $\mu_1 - \mu_2$.

(7) عرف وبرهن فترة $(1-\alpha)100\%$ ثقة للفرق بين الوسطين الحسابين $\mu_1 - \mu_2$ إذا كان حجم العينات أقل من 50 والتباينات مجهولة.

(8) أريد مقارنة نسبة المدخنين في مدينتي بغداد والبصرة. ولهذا الغرض أخذت عينة من سكنة مدينة بغداد حجمها 500 شخص فوجد أن منهم 250 مدخنا. كما أخذت عينة من سكان مدينة البصرة حجمها 900 فوجد أن منهم 540 مدخنا. المطلوب حساب فترة ثقة 90% للفرق بين نسبتي المدخنين في بغداد والبصرة.

(9) أخذت عينتان من توزيعين طبيعيين كانت لهما البيانات التالية:

العينة الأولى	$n = 10$	$\bar{X} = 50$	$S_1^2 = 100$
العينة الثانية	$m = 8$	$\bar{Y} = 40$	$S_2^2 = 120$

أوجد فترة ثقة للفرق بين الوسطين $\mu_1 - \mu_2$ وذلك عندما:

(1) يكون التوزيعان لهما نفس التباين (أي $\sigma_1^2 = \sigma_2^2 = \sigma^2$).

(2) يكون التوزيعان بتباينين مختلفين ($\sigma_1^2 \neq \sigma_2^2$).

(10) أخذت عينتان حجم الأولى 10 وحجم الثانية 15 من توزيعين طبيعيين تبايناتهما مجهولة، وقد وجد أن $S_1^2 = 29$ ، $S_2^2 = 24$. أوجد فترة ثقة 90% للنسبة σ_1^2 / σ_2^2 .

(11) سحبت عينة من 25 مفردة، حسب متوسطها وانحرافها المعياري فكانا 60 ,9 على التوالي. والمطلوب:

(1) فترة 90% ثقة لوسط المجتمع الذي سحبت منه العينة.

(2) فترة 95% ثقة للتباين σ^2 .

(12) العينات الآتية:

(أ) 11, 15, 14, 2, 11, 5, 20, 14, 17

(ب) 1, 4, 8, 10, 12

مأخوذة من مجتمعات طبيعية. أوجد لكل واحدة منهن أحسن التقديرات لكل من: σ ، σ^2 ، μ .

(13) إذا سحبت عينات من مجتمعات طبيعية تبايناتها معلومة، المطلوب إيجاد:

(1) درجة الثقة المستخدمة إذا علم أن $n = 36$ ، $\sigma = 8$ وأن المدى الكلي لفترة الثقة حول المتوسط يساوي 3.28 وحدة.

(2) حجم العينة إذا علم أن $\sigma = 10$ وأن فترة الثقة حول المتوسط عند درجة ثقة 95% هي 16.2 إلى 21.8 وحدة.

(3) التباين عندما تكون $n = 100$ وأن فترة الثقة حول المتوسط عند درجة ثقة 98% بمدى يساوي 23.26 وحدة.

(14) فرق بين: (أ) تقدير النقطة وتقدير الفترة. (ب) الكفاءة والكفاية.

(15) إذا كانت دالة كثافة الاحتمال للمتغير العشوائي المستمر X هي:

$$f(x,\theta)=\frac{nx^{n-1}}{\theta^{n}}, \qquad 0<x<\theta$$

فاثبت أن $\hat{\theta}$ حيث: $\hat{\theta}=(1+\frac{1}{n})x$ ، تقدير غير متحيز للمعلمة θ .

(16) كان مجموع مربعات القيم لعينة حجمها 10 مسحوبة من مجتمع طبيعي $N(\mu,\sigma^2)$ يساوي 58 وكان $\sum X=20$ ، المطلوب:

(1) فترة ثقة 99% لـ $(0.5-\mu)$.

(2) فترة ثقة 90% لـ $\sigma / 2$.

(17) إذا كان $S^2=\frac{1}{r}\sum(X-\overline{X})^2$ حيث r تمثل درجات الحرية فان المطلوب إثبات أن:

$$Var(S^2)=\frac{2\sigma^4}{r}$$

(18) أفرض أن X_1, X_2 متغيرات عشوائية مستقلة تتبع توزيع بواسون. المطلوب مقارنة كفاءة تقدير متوسط هذه المتغيرات مع كفاءة التقدير $X_1+2X_2/3$.

(19) من العلوم أن S^2 ، حيث $S^2=\frac{1}{n}\sum(X-\overline{X})^2$ ، تقدير متحيز للمعلمة σ^2 . احسب مقدار التحيز.

(20) من مجتمع يتكون من المفردتين 7 , 3 سحبت كل العينات من الحجم $n = 3$ بالإرجاع. المطلوب إيجاد التوزيع الاحتمالي للوسيط والتحقق من تحيز أو عدم تحيز الوسيط بالنسبة للمعلمة μ .

(21) ما هو الفرق بين خطأ المعاينة والدقة؟

.

الفصل الثالث

اختبارات الفروض

(1.3) مقدمة.

(2.3) أنواع الخطأ.

(3.3) الاختبار ذو الذيل الواحد أو الذيلين.

اختبارات الفروض
Test of Hypothesis

(1.3) مقدمة (*Introduction*)

تظهر وفي حالات كثيرة مشكلة مدى تمثيل العينة المسحوبة عشوائيا من المجتمع الذي سحبت منه، بمعنى معرفة مدى انتماء العينة للمجتمع المسحوبة منه والذي له صفاته الخاصة. فمن المعروف انه ليس بالضرورة لأي إحصائية من إحصائيات العينة أن تساوي معلمة المجتمع المناظرة. فإحصائية العينة $\overline{X}$ قد تكون أكبر أو أصغر من معلمة المجتمع μ ، وفي حالات نادرة جدا تكون $\overline{X}$ مساوية لـ μ حسابيا.

على انه إحصائيا قد يوجد فرق بين الإحصائية $\overline{X}$ والمعلمة μ (سواء بالزيادة أو النقص)، غير أننا يمكن أن نتجاهل هذا الفرق إحصائيا باعتباره فرق ظاهري وغير حقيقي وليس له خطورة على مدى تمثيل العينة للمجتمع. وعليه يمكن اعتبار أن متوسط العينة لا يختلف إحصائيا عن متوسط المجتمع وان العينة مسحوبة فعلا من المجتمع المعلوم، وفي هذه الحالة نعتبر أن : $\overline{X} = \mu$ إحصائيا.

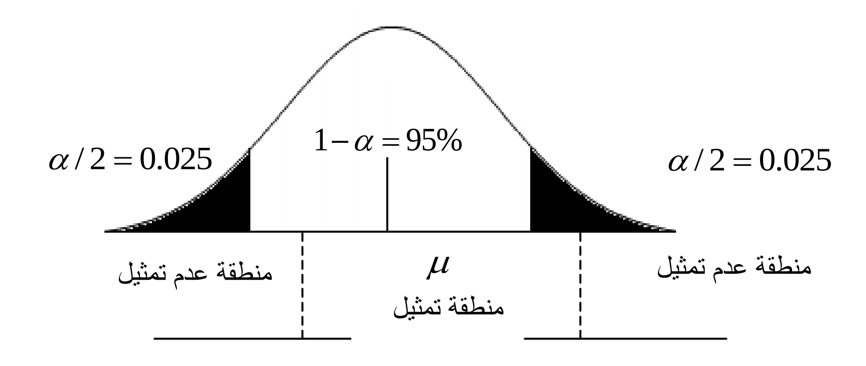
ولكن وفي بعض الحالات الأخرى لا يمكن تجاهل هذا الفرق بين $\overline{X}$ ، μ حيث نعتبره فرقا حقيقيا وجوهريا وهو يمثل خطورة على مدى تمثيل العينة للمجتمع، وهنا يكون متوسط العينة مختلفا إحصائيا عن متوسط المجتمع وان العينة لا تمثل هذا المجتمع المعلوم، وفي هذه الحالة نعتبر أن $\overline{X} \neq \mu$ إحصائيا. وفي الحقيقة فان

مسالة تحديد ما إذا كان الفرق بين $\overline{X}$ ، μ معنويا أو غير معنوي ليس متروكا للتقدير الشخصي للباحث، ولكنه يبنى على الخصائص الأساسية للتوزيع العيني للمتوسطات المشار إليها سابقا، فقد ذكرنا أن من خصائص التوزيع العيني للمتوسطات الحسابية أن 95% من المتوسطات الحسابية لكل العينات الممكنة المسحوبة من المجتمع تقع في الفترة:

$$\mu \pm 1.96(\frac{\sigma}{\sqrt{n}})$$

أي بين متوسط المجتمع μ وضعف الخطأ المعياري تقريبا من اليسار واليمين بدرجة ثقة قدرها 95% . وتأسيسا على ذلك إذا كانت قيمة الوسط للعينة $\overline{X}$ تقع داخل الحدود فانه يمكن اعتبار أن $\overline{X}$ لا تختلف معنويا عن μ ، أي أن الفرق بين $\overline{X}$ ، μ غير معنوي بمستوى $(1-\alpha)100\%$ رغم اختلاف متوسط العينة عن متوسط المجتمع ظاهريا. والشكل التالي يوضح لنا هذه الحدود عند مستوى معنوي 5%.

منطقة تمثيل وعدم تمثيل الإحصائية $\overline{X}$ للمعلمة μ عند $\alpha = 5\%$



شكل (رقم 8)

ومما تقدم عرضه فسوف يكون هدفنا هو البحث عن أسلوب إحصائي عن طريقه نستطيع أن نحكم على الفروق بين التقدير والمعلمة المجهولة، بمعنى هل أن هذه الفروق هي فروق معنوية أو غير معنوية حيث تتوقف جميع قراراتنا على نوع الفرق. ومن هنا تأتي أهمية اختبارات الفروض كواحدة من أهم التطبيقات لعلم الإحصاء في الحياة العملية. إذ انه باستخدام نظرية الاحتمالات وخصائص التوزيعات العينية يمكننا أن نتخذ قرارا برفض أو قبول فرض معين أو مجموعة من الفروض.

وقد تتناول الاختبارات التي ستناقشها الدراسة فروضا متعلقة بمعالم المجتمع مثل الفروض الخاصة بالمتوسط أو بالنسبة أو التباين ... الخ، أو قد لا تتناول دراسة فروض خاصة بمعالم المجتمع المجهولة مثل الفروض الخاصة بالعلاقة بين الحالة التعليمية والحالة الاجتماعية مثلا أو العلاقة بين لون الشعر ولون العين ... الخ. ودراسة الفروض في الحالة الأولى يتطلب معرفة التوزيع الاحتمالي للمجتمع ويطلق على الاختبارات في هذه الحالة بالاختبارات المعلمية (*Parametric tests*)، على عكس دراسة الفروض في الحالة الثانية والتي لا يتطلب فيها معرفة نوع التوزيع الاحتمالي للمجتمع، ويطلق على هذا النوع من الاختبارات بالاختبارات اللامعلمية (*Non-Parametric tests*).

وبشكل عام فان فكرة اختبارات الفروض الإحصائية تعتمد على وضع فرض معين بخصوص المشكلة موضع الدراسة والفرض البديل له كخطوة أولى للوصول إلى القرار مع اقتران هذا الفرض بمستوى معنوية معين نتيجة لخطا الصدفة أو العشوائية الناتج من الاعتماد على بيانات العينة حيث يتوزع هذا الخطأ بطريقة احتمالية. ومن بيانات العينة يمكننا تحديد التوزيع الاحتمالي الذي تتبعه عملية الاختبارالاحصائي، أي الصيغة الرياضية الملائمة للاختبار اللازم للتطبيق وأخيرا أخذ القرار المناسب. ولكي نتعرف على اختبارات الفروض كموضوع جديد يجب أن نتعرف على المراحل المذكورة أعلاه بالتفصيل في أدناه.

أولا: وضع الفروض

أن وضع فرض معين يقوم على أساسه اتخاذ القرار هو نقطة البداية في الوصول إلى القرار السليم بصدد المشكلة موضع الدراسة. وعند وضع هذا الفرض فانه في الواقع ينطوي على وضع فرض آخر بديل، ليكون لدينا فرضان لمشكلة واحدة وهما فرضان مانعان بالتبادل، بمعنى إما أن نرفض الفرض الأول وبالتالي نقبل الثاني أو العكس، وهذان الفرضان هما:

(أ) **الفرض العدم** $(Null\ Hypothesis)$: وهو الفرض الأول الذي سوف نعطي له الرمز H_0 . ويقصد به عدم وجود اختلاف أو فرق كبير جوهري أو حقيقي أو معنوي بين إحصائية العينة ومعلمة المجتمع، وان هذه العينة المسحوبة من المجتمع ممثلة لهذا المجتمع وان وجد أي اختلاف إنما يمكن إرجاعه للصدفة والعشوائية وهو اختلاف غير مقصود.

(ب) **الفرض البديل** $(Alternative\ Hypothesis)$: وهو الفرض الثاني الذي نعطي له الرمز H_1، ويقصد به وجود اختلاف جوهري بين إحصائية العينة ومعلمة المجتمع، وهذا الفرق حقيقي ومعنوي لا يمكن إرجاعه لمحض الصدفة وان بيانات العينة تؤكد وتؤيد صحة هذا الفرق.

فإذا كان المطلوب مثلا اختبار أن معلمة المجتمع μ تساوي قيمة محددة افتراضية لمعلمة المجتمع ولتكن μ_0 لا تختلف عنها معنويا، فإننا نضع الفروض على الشكل التالي:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

أما إذا كان المطلوب اختبار ما إذا كانت معلمة المجتمع لا تزيد عن قيمة معينة μ_0 فإننا نضع الفروض على الشكل التالي:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

وأخيرا إذا كان المطلوب اختبار ما إذا كانت معلمة المجتمع لا تقل عن حد معين μ_0 ، فإن الفروض تكون على الشكل الآتي:

$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

وهذا يعني أن صور الفرض العدم أو الفرض البديل عديدة وتتوقف على نوعية المشكلة موضع الحل.

ثانيا: اختيار مستوى المعنوية *(Selection of the level of Significance)*

لأننا نستخدم بيانات العينة في حساب الإحصائية فإننا نقع عند إجراء الاختبار الإحصائي في الأخطاء العشوائية والتي تخضع أساسا لأحد التوزيعات الاحتمالية المعروفة، لذلك لابد من ربط قبول أو رفض الفروض الإحصائية بدرجة ثقة معينة أو مستوى معنوية معين، وهي التي تحدد درجة تحملنا لأخطاء معينة عند اتخاذ القرار. ولقد جرى أسلوب العمل الإحصائي في مجال الاختبارات الإحصائية على أخذ مستوى المعنوية α يساوي 5% أو 1%. وعندما تكون $\alpha = 0.05$ فهذا معناه أنه في مائة حالة مماثلة يمكن لنا رفض الفرض الأول في 5 حالات فقط على الرغم من أنه فرض صحيح، وهي نسبة صغيرة بالطبع. أما النسبة الكبيرة $1-\alpha = 95\%$ فهي تمثل باقي الحالات التي نقبل فيها الفرض الأول وهو صحيح.

تعريف (1): مستوى المعنوية

وهي درجة الاحتمال التي نرفض بها فرضية العدم H_0 عندما تكون صحيحة ويرمز لها بالرمز α .

(2.3) أنواع الخطأ (*Types of Error*)

انتهينا إلى أننا سوف نتخذ قرارا بناء على المعلومات المتحصل عليها من العينة، ألأمر الذي يعرضنا إلى ارتكاب بعض الأخطاء. وما يهمنا هو التعرف على هذه الأخطاء وعلى كيفية حساب احتمال وقوعها، وهذا ما يساعدنا على حساب احتمال اتخاذ قرار

خاطئ وبالتالي تقدير كمية الأخطاء (*Risks*) التي تواجهنا باستخدامنا لهذا القرار. وبصفة عامة يواجه الباحث عند إجراء الاختبار الإحصائي نوعين من الأخطاء **وهما:**

(أ) الخطأ من النوع الأول (*Type I error*)

وهو الخطأ الناتج عن رفضنا لفرض العدم H_0 على الرغم من انه فرض صحيح، ويرمز له بالرمز α، حيث تمثل α احتمال رفض الفرضية الأولى H_0 بالرغم من صحتها، أي أن:

$$\begin{aligned}\alpha &= p(Type\ I\ error)\\ &= p(reject\ H_0 \mid H_0\ is\ true)\end{aligned}$$

وبذلك فان الخطأ من النوع الأول عبارة عن مستوى المعنوية α، وقرار رفضنا للفرض H_0 على الرغم من صحته إنما يعني قبولنا للفرض البديل H_1 على الرغم من عدم صحته.

أما احتمال قبولنا لفرض العدم عندما يكون صحيحا فهو قرار صائب ويعرف بـ $1-\alpha$ حيث:

$$\begin{aligned}1-\alpha &= 1-p(reject\ H_0 \mid H_0\ is\ true)\\ &= p(accept\ H_0 \mid H_0\ is\ true)\end{aligned}$$

(ب) الخطأ من النوع الثاني (*Type II error*)

وهو الخطأ الناتج عن قبولنا لفرض العدم H_0 على الرغم من انه غير صحيح، ويرمز له بالرمز β والتي تمثل احتمال قبولنا للفرضية الأولى على الرغم من عدم صحتها، أي أن:

$$\begin{aligned}\beta &= p(Type\ II\ error)\\ &= p(accept\ H_0 \mid H_0\ is\ false)\end{aligned}$$

وهذا معناه إننا نرفض H_1 على الرغم من انه صحيح. أما $1-\beta$ فهي تمثل احتمال رفضنا للفرض العدم عندما يكون هو الفرض الخاطئ، ويسمى هذا الاحتمال بقوة الاختبار (*Power of the test*) وتعرف كالتالي:

$$1-\beta = 1-p(accept\ H_0 \mid H_0\ is\ false)$$
$$= p(reject\ H_0 \mid H_0\ is\ false)$$

ومن هنا يتضح أنه كلما قلت β زادت قوة الاختبار. وبصفة عامة فإنه يتعين على الإحصائي والذي يستخدم أي نوع من أنواع الاختبارات الإحصائية أن يراعي عنصر التوازن بين وقوعه في الخطأ من النوع الأول والخطأ من النوع الثاني فلا يغلب خطأ على آخر ويعلم أن الخبرة وممارسة العمل الإحصائي في هذا المجال لها دورها. فأخذ قيمة كبيرة لـ α من شأنه أن يقلل من قيمة β، كما أن تحديد قيمة كبيرة للخطأ β من شأنه تقليل قيمة α، مع العلم أن قيمة α تحدد أولا وقبل البدء في عمل الاختبار على خلاف تحديد قيمة β والتي تحدد بعد تحديد منطقة القبول والرفض للفرضيات الموضوعة. ويمكن تلخيص النوعين السابقين في الجدول التالي.

جدول (رقم 21)

	H_0 ***is true***	H_0 ***is false***
reject H_0	α	$1-\beta$
accept H_0	1 - α	β

وتجدر الملاحظة بان هناك علاقة بين الخطأ من النوع ألأول α والخطأ من النوع الثاني β **تتلخص بالآتي:**

1. أن نقصان أحدهما يزيد من ألأخرى.
2. إن زيادة حجم العينة يقلل من احتمال كل من α، β.
3. إن α تحسب على أساس قيمة فرض العدم بينما تحسب β على أساس قيمة الفرض البديل.

ويمكن حساب احتمال وقوع أي خطأ من الخطأين السابقين كما يلي:

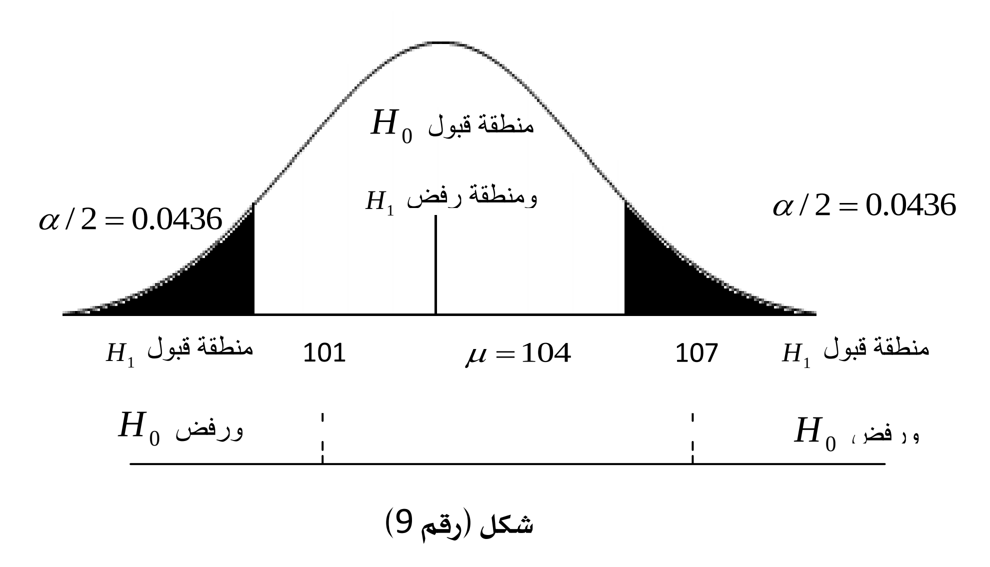
أولا: حساب احتمال الخطأ من النوع الأول

وفي هذه الحالة يكون المطلوب حساب احتمال قبول الفرض البديل H_1 في حين أن الفرض العدم H_0 صحيح. ولحساب هذا الاحتمال نورد المثال التالي.

- **مثال (35):** آلة تستخدم لملئ أكياس معينة بطريقة اتوماتيكية بمتوسط 104 غم علما بأن الانحراف المعياري للمجتمع هو 8.75 غم. سحبت عينة من 25 وحدة من ذلك المجتمع واختبرت وكررت هذه العملية كل ساعة لتحديد إذا كانت الآلة تنتج حسب المواصفات أم لا. افرض أن الإنتاج سوف يستمر إذا كان $101 \leq \overline{X} \leq 107$ ، وانه سوف يتوقف إذا كان $\overline{X} < 101$ أو $\overline{X} > 107$. فان المطلوب حساب قيمة احتمال الخطأ من النوع الأول α مع احتمال اتخاذ قرار صحيح.

الحل:

إذا كانت قواعد اتخاذ القرار هي رفض الفرض العدم عندما تكون $\overline{X} > 107$ أو $\overline{X} < 101$ ، فإن احتمال الخطأ من النوع الأول هو المناطق المظللة من الشكل التالي.



شكل (رقم 9)

ولحساب مساحة الأجزاء المظللة فنحن نعلم أن $Z_1 = \frac{(\overline{X}_1 - \mu)\sqrt{n}}{\sigma}$، حيث $\overline{X}_1$ هي القيمة الحرجة للمتوسط μ، تتبع التوزيع المعتدل المعياري بمتوسط μ وانحراف معياري $\frac{\sigma}{\sqrt{n}}$. إذن القيمة المعيارية Z_1 للقيمة الحرجة 101 هي:

$$Z_1 = \frac{\overline{X}_1 - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{101 - 104}{\frac{8.75}{\sqrt{25}}} = -1.71$$

القيمة المعيارية Z_2 للقيمة الحرجة 107 هي:

$$Z_2 = \frac{\overline{X}_2 - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{107 - 104}{\frac{8.75}{\sqrt{25}}} = -1.71$$

وباستخدام جداول المنحنى المعياري في الملحق نجد أن احتمال الخطأ من النوع الأول α يساوي:

$$\begin{aligned} \alpha &= p(reject\ H_0 \mid H_0\ is\ true) \\ &= 1 - 2p(0 \le Z \le 1.71) \\ &= 1 - 2(0.4564) \\ &= 0.0872 \end{aligned}$$

أما احتمال اتخاذ قرار صحيح فهو يساوي:

$$\begin{aligned} 1 - \alpha &= p(reject\ H_0 \mid H_0\ is\ true) \\ &= 1 - 0.0872 \\ &= 0.9128 \end{aligned}$$

ومن هذا المثال يتضح أنه بإمكان الباحث أن يتحكم في الخطأ من النوع الأول وذلك بتحريك القيم الحرجة للمتوسط مبتعدا عن أو مقتربا من متوسط التوزيع العيني.

ويمكن كذلك إيجاد القيمة الحرجة التي تعطي خطأ من النوع الأول مقداره $\alpha/2$ من الطرفين، وذلك بأن نوجد القيمة المعيارية من جدول المنحنى الطبيعي المعياري والتي تعطي احتمالا مقداره $\alpha/2$ في الذيل ولتكن $\pm Z_{\alpha/2}$، وحيث أن:

$$Z_{\alpha/2} = \frac{\overline{X}_1 - \mu}{\frac{\sigma}{\sqrt{n}}}$$

فإن القيمة الحرجة التي على يمين المتوسط هي:

$$\overline{X}_1 = (Z_{\alpha/2})(\frac{\sigma}{\sqrt{n}}) + \mu$$

أما القيمة الحرجة التي على يسار المتوسط فإننا نحصل عليها بتغيير إشارة $Z_{\alpha/2}$، وبالتالي فإن:

$$\overline{X}_2 = \mu - (Z_{\alpha/2})(\frac{\sigma}{\sqrt{n}})$$

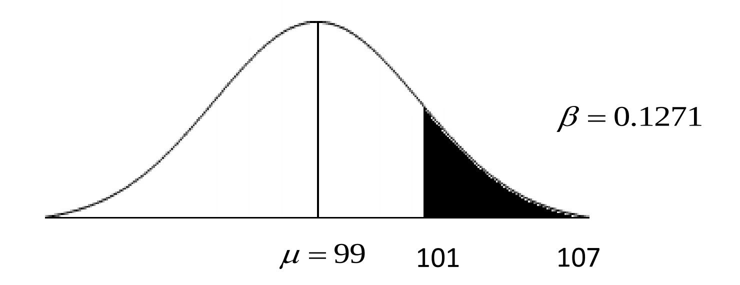
ثانيا: حساب احتمال الخطأ من النوع الثاني

وهو احتمال قبول الفرض العدم في حين أن الفرض البديل هو الصحيح، بمعنى أنه قد تم قبول الفرض العدم وهو غير صحيح. والمثال التالي يوضح طريقة الحساب لهذا الاحتمال.

- **مثال (36):** في مثال (35)، المطلوب حساب الخطأ من النوع الثاني β عندما يكون متوسط المجتمع يساوي 99 و 109.

الحل:

(أ) من المجتمع الذي متوسطه 99 يكون المطلوب إيجاد المساحة المحصورة ما بين 101,107 كما هو واضح من الشكل التالي:



شكل (رقم 10)

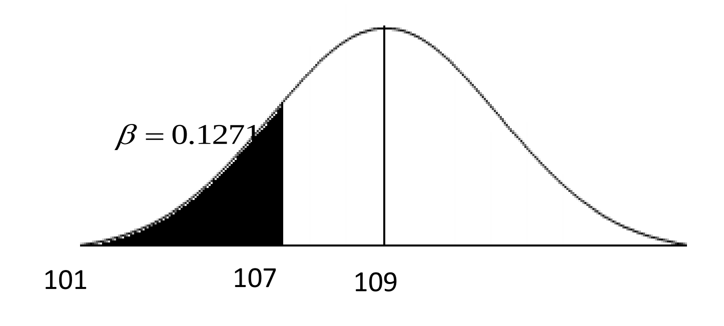
واضح من الرسم أن المساحة المطلوبة هي تقريبا كل المساحة في الذيل الأيمن للتوزيع وتحسب باستخدام جداول المنحنى المعتدل المعياري كما يلي:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{101 - 99}{\frac{8.75}{\sqrt{25}}} = 1.14$$

وبالتالي فان قيمة β هي:

$$\beta = p(accept\ H_0 \mid H_0\ is\ false)$$
$$= p(Z \geq 0) - p(0 \leq Z \leq 1.14) = 0.5 - 0.3729 = 0.1271$$

(ب) إذا كان متوسط المجتمع يساوي 109 فإن الاحتمال المطلوب سوف يوضحه الشكل التالي:



شكل (رقم 11)

وهنا نلاحظ أن المساحة المطلوبة هي كل المساحة في الذيل الأيسر للتوزيع تقريبا ليكون حسابها على الشكل التالي:

$$\beta = p(\overline{X} \le 109) - p(107 \le \overline{X} \le 109)$$
$$= p(Z \le 0) - p(-1.14 \le Z \le 0)$$
$$= p(Z \ge 0) - p(0 \le Z \le 1.14) = 0.5 - 0.3729 = 0.1271$$

وكقاعدة عامة نجد الحالتين الآتيتين:

الحالة الأولى: إذا كان الفرض العدم صحيحا فإن:

(أ) احتمال القرار الصحيح هو المساحة المحصورة تحت التوزيع العيني النظري للمتوسط بين القيمتين الحرجتين للمتوسط $\overline{X}_1, \overline{X}_2$.

(ب) احتمال الخطأ من النوع الأول هو تلك المساحة تحت التوزيع العيني بعد وقبل النقطتين الحرجتين $\overline{X}_1, \overline{X}_2$ على التوالي.

الحالة الثانية: إذا كان الفرض البديل صحيحا فإن:

(أ) احتمال القرار الصحيح هو تلك المساحة تحت التوزيع العيني النظري قبل وبعد النقطتين الحرجتين $\overline{X}_1, \overline{X}_2$ على التوالي.

(ب) احتمال الخطأ من النوع الثاني هو تلك المساحة تحت التوزيع العيني المحصورة ما بين القيمتين الحرجتين $\overline{X}_1, \overline{X}_2$.

ثالثا: تحديد نوع التوزيع الاحتمالي (المختبر الإحصائي)

ويتم ذلك من خلال تحديد نوعية إحصائية العينة ومدى انتمائها إلى توزيع احتمالي معين من خلال البيانات أو المشاهدات المأخوذة من العينة لغرض اختبار فرضية المجتمع. وبصفة عامة فإن التوزيعات الاحتمالية يوجد منها ما يناسب كل اختبار على حده. فعند اختبار العلاقة بين متوسط عينة ومتوسط مجتمع وكان حجم العينة كبيرا فإن التوزيع الاحتمالي النظري المناسب هو التوزيع الطبيعي في إيجاد قيمة Z المعيارية وهكذا.

تعريف (2): المختبر الإحصائي: وهو عبارة عن متغير عشوائي له توزيع احتمالي معلوم يصف العلاقة بين القيم النظرية للمجتمع والقيم المحسوبة من العينة.

رابعا: اتخاذ القرار بالرفض أو القبول

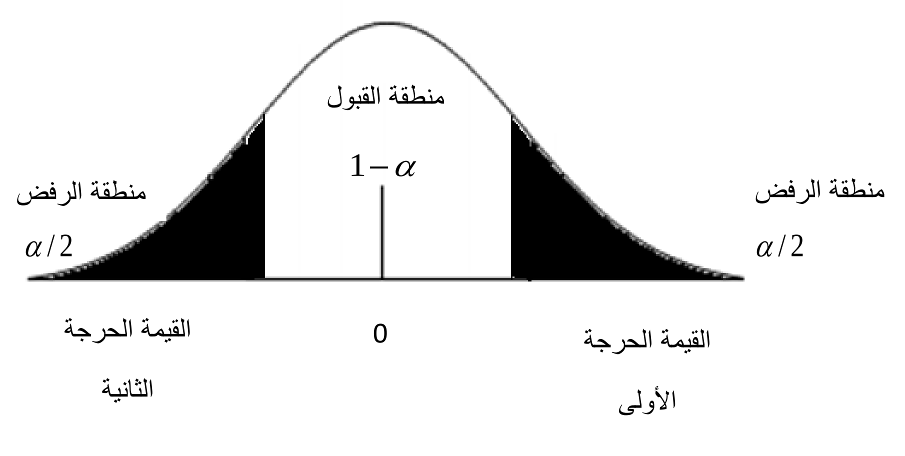
بعد أن تحول الإحصائية إلى شكل معياري وذلك عن طريق قسمة الفرق بين الإحصائية (المستخرجة من المشاهدات) وفرضية المجتمع μ على الخطأ المعياري لكل حالة، فإننا نقارن هذه القيم المشاهدة بالقيم النظرية المستخرجة من الجداول المعدة للتوزيعات الاحتمالية (مثل جدول توزيع Z، جدول توزيع t، جدول توزيع χ^2 أو جدول توزيع F ...الخ) لاتخاذ القرار برفض أو قبول الفرض العدم.

(3.3) الاختبار ذو الذيل الواحد أو الذيلين *One and Two Tail Tests*

يطلق على الاختبار الذي يكون فيه منطقتان للرفض ومنطقة واحدة للقبول بالاختبار ذو الذيلين أو الاختبار من الطرفين، ويكون فيه فرض العدم مقابل الفرض البديل على الشكل التالي:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

حيث μ_0 قيمة معينة معلومة. والشكل التالي يوضح ذلك.



شكل (رقم 12)

وهنا إذا وقعت القيمة المحسوبة داخل حدود منطقة القبول *(Acceptance Region)* فإننا نقبل فرضية العدم الأولى ونرفض فرضية البديل الثانية في ضل مستوى المعنوية α المأخوذة سلفا قبل إجراء الاختبار. أما إذا وقعت القيمة المحسوبة خارج حدود منطقة القبول، أي وقعت في منطقة الرفض *(Rejection Region)* بما في ذلك القيمتان الحرجتان نفسهما، فإننا نرفض H_0 ونقبل H_1 عند α .

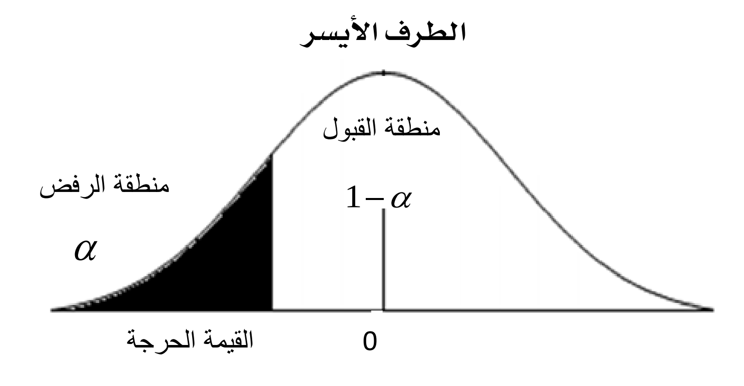
أما إذا كان للاختبار منطقة رفض واحدة فإنه يطلق عليه اختبار ذو ذيل واحد أو الاختبار من الطرف الواحد، **وهنا نلاحظ الحالتين الآتيتين:**

حالة اختبار الطرف الأيسر: وفيها يكون الفرض العدم مقابل الفرض البديل كما يلي:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu < \mu_0$$

والشكل التالي يوضح هذا النوع من الاختبار.

منطقة القبول ومنطقة الرفض في اختبار الطرف الأيسر



شكل (رقم 13)

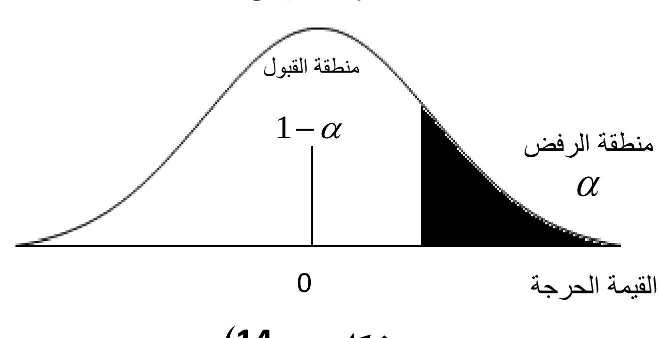
وهنا إذا وقعت القيمة المحسوبة على يمين القيمة الحرجة المستخرجة من جدول التوزيع الاحتمالي، أي وقعت في منطقة القبول فإننا نقبل H_0 ونرفض H_1 عند مستوى المعنوية المأخوذة. أما إذا وقعت القيمة المحسوبة على يسار القيمة الحرجة (بما

في ذلك نفس القيمة الحرجة المستخرجة من الجدول) أي أنها وقعت في منطقة الرفض، فإننا نرفض H_0 ونقبل H_1 .

حالة اختبار الطرف الأيمن: وفي هذه الحالة فإن الفروض تكون على الشكل التالي:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

منطقة القبول ومنطقة الرفض في اختبار الطرف الأيمن



شكل رقم 14)

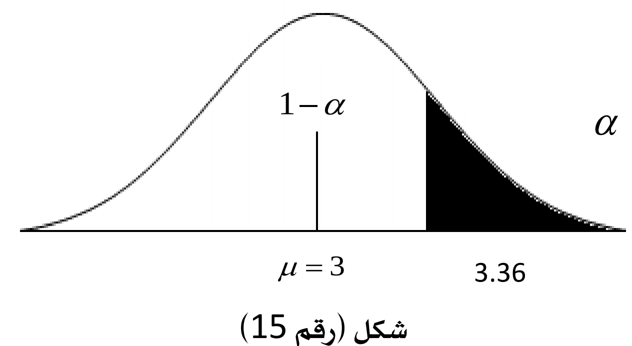
وهنا يرفض الفرض العدم H_0 عند مستوى المعنوية إذا ظهرت أن القيمة المحسوبة أكبر من أو تساوي القيمة الجدولية، لأن هذا يعني أن القيمة المحسوبة تقع في منطقة الرفض. أما إذا كانت القيمة المحسوبة أقل من القيمة الجدولية فإننا نقبل H_0 ونرفض H_1 حيث لا يوجد ما يدعم أو يؤيد رفض هذا الفرض.

- **مثال (37)**: أفرض أن مجتمعا متوسطه يساوي 3 وأن انحرافه المعياري يساوي 1.5 سحبت عينة حجمها 81 مفردة وكان متوسط العينة يساوي 3.36، المطلوب حساب احتمال الخطأ من النوع الأول واحتمال اتخاذ قرار صحيح إذا كان:

$$H_0 : \mu = 3$$
$$H_1 : \mu > 3$$

الحل:

الشكل التالي يوضح التمثيل البياني لمثالنا أعلاه.



شكل (رقم 15)

وبذلك تكون α هي تلك المساحة الموجودة في الطرف الأيمن للتوزيع بعد القيمة الحرجة 3.36. وباستخراج القيمة المعيارية Z حيث:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.36 - 3}{\frac{1.5}{\sqrt{81}}} = 2.16$$

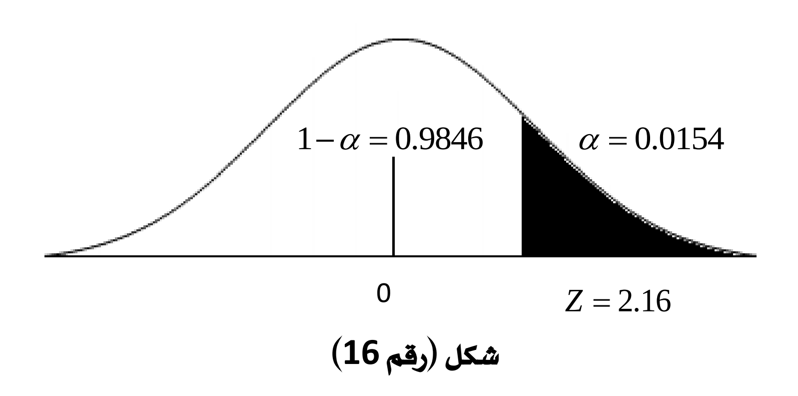
فإنه يمكن إيجاد الاحتمال α باستخدام جدول التوزيع الطبيعي المعياري في الملحق كما يلي:

$$\begin{aligned}
\alpha &= p(Type\ I\ Error) \\
&= p(reject\ H_0 \mid H_0\ is\ true) \\
&= p(\overline{X} \geq 3.36) \\
&= p(Z \geq 2.16) \\
&= p(Z \geq 0) - p(0 \leq Z \leq 2.16) = 0.5 - 0.4846 = 0.0154
\end{aligned}$$

أما احتمال اتخاذ قرار سليم فهو:

$$\begin{aligned} 1-\alpha &= 1-p(reject\ H_0 | H_0 is\ true) \\ &= 1-0.0154 = 0.9846 \\ &= p(accept\ H_0 | H_0\ is\ true) \end{aligned}$$

والشكل التالي يوضح كلا من α ، $1-\alpha$.



شكل (رقم 16)

تمـــارين

(1) عرف كلا من:

(1) الخطأ من النوع الأول والخطأ من النوع الثاني.

(2) الاختبار من طرف واحد والاختبار من طرفين.

(2) ما هي العلاقة بين الخطأ من النوع الأول والخطأ من النوع الثاني؟ وهل تتأثر هذه الأنواع من الأخطاء بزيادة حجم العينة؟

(3) عرف ما يأتي: الفرضية الإحصائية، مستوى المعنوية، المختبر الإحصائي.

(4) عدد خطوات اختبار الفرضيات.

(5) استخدمت ماكينة لملئ أكياس من الطحين بمتوسط قدره 104 كغم للكيس الواحد حيث كان الانحراف المعياري للمجتمع هو 4 كغم. فإذا سحبت عينة تتكون من 25 وحدة من ذلك المجتمع، واختبرت وكررت العملية كل ساعة لتحديد إذا كانت الآلة تنتج حسب المواصفات أم لا. والمطلوب باستخدام احتمال 95% حساب القيم الحرجة للمتوسط لتحديد متى تكون الآلة مضبوطة ومتى تكون غير ذلك.

(6) سحبت عينة حجمها (36) من مجتمع متوسطه (6) فوجد أن: $\sum(X-5)=-36, \sum X^2 = 1008$، والمطلوب حساب احتمال الخطأ من النوع الأول α واحتمال اتخاذ قرار صحيح $1-\alpha$.

الفصل الرابع

الاختبارات المعلمية

الاختبارات المعلمية

Parametric Test

(1.4) مقدمة (*Introduction*)

بعد أن تحدثنا عن اختبارات الفروض والى بعض التعريفات اللازمة لها نريد أن نتحدث الآن عن الاختبارات المعلمية وهي الاختبارات التي تتعلق بمعالم المجتمع المجهولة. ولإجراء أي اختبار إحصائي يجب أن نمر بخطوات معينة محددة تعرضنا لها في الباب السابق تفصيلا وهي لا تتغير بتغير الاختبار **ويمكن تلخيصها بالآتي:**

الخطوة الأولى: وهي صياغة الفرض العدم الذي سيجرى على أساسه الاختبار.

الخطوة الثانية: وهي صياغة الفرض البديل والذي سيقبل عند رفض الفرض العدم.

الخطوة الثالثة: تحديد الخطأ من النوع الأول α الذي يمثل مقدار الخطر الذي سوف نتعرض له عند اتخاذ القرار.

الخطوة الرابعة: باستخدام نظرية الإحصاء نبحث عن توزيع عيني يتبع له التقدير، وهو التوزيع الذي على أساسه يتم رفض أو قبول فرض العدم.

الخطوة الخامسة: استخدام المعلومات التي تم تجميعها من العينة لاتخاذ القرار الذي يقضي إما برفض أو قبول فرض العدم.

وما يتضح من الخطوات السابقة هو أن الخطوات الثلاثة الأولى تحدد من قبل الباحث نفسه. أما الخطوة الرابعة فلقد تم إيجاد توزيعات احتمالية للمشاكل المختلفة بالبرهان الرياضي الذي سوف لن نتعرض له بل سنتعامل مع كل مشكلة على حدة مفترضين أن لها توزيعا عينيا معينا نذكره ونذكر كيفية استخدامه في

قبول أو رفض فرض العدم بناء على المعلومات المتاحة من العينة وهذه هي الخطوة الخامسة. وسوف نتعرض في هذا الباب إلى الاختبارات المعلمية بتقسيمها إلى اختبارات خاصة بعينة واحدة ثم إلى اختبارات خاصة بعينتين.

(2.4)اختبارات تتعلق بعينة واحدة

Tests Concerning one Sample

إذا تم اختيار عينة واحدة من المجتمع وأردنا اختبار بعض الفروض التي تتعلق بمعالم المجتمع المجهولة بناءا على المعلومات المتاحة من هذه العينة، فانه يمكن إجراء بعض الاختبارات كاختبار المتوسط النسبة، التباين، معامل الارتباط ... الخ. وهذا ما سنتعرض له في الفقرات التالية.

(1.2.4) اختبار المتوسط في العينة (Tests for Sample Mean)

إذا سحبنا عينة واحدة من مجتمع معين وأردنا أن نختبر هل هذه العينة مسحوبة من مجتمع له متوسط معين أم لا، أو بصفة عامة إذا أردنا استخدام المعلومات المتاحة من العينة لاختبار بعض الفروض حول متوسط المجتمع، فيجب علينا أن نفرق بين حالتين.

أولا: إذا كان تباين المجتمع معلوما

هنا إذا ما أردنا اختبار العلاقة بين متوسط العينة $\bar{X}$ ومتوسط المجتمع μ المسحوبة منه العينة لمعرفة ما إذا كان الفرق ($\bar{X}$ - μ) معنويا أو غير معنوي بدرجة ثقة معينة وكان التباين σ^2

معلوما، فانه تم إثبات أن المتغير Z :

$$Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$$

يتوزع توزيعا معتدلا معياريا. وهذا هو التوزيع العيني الذي عن طريقه سوف نحدد مناطق الرفض والقبول علما بأن μ_0 قيمة محددة معلومة. ولإجراء الاختبار نجري الخطوات الآتية:

(1) تحديد الفروض ومنطقة كل فرض:

الفرض الأول: هو فرض العدم H_0 والذي يعني عدم وجود فرق معنوي بين $\overline{X}$ ، μ ، حيث:

$$H_0 : \mu = \mu_0$$

أي أن الفرق $\mu - \mu_0$ = صفرا وأن العينة تنتمي إلى المجتمع (μ, σ^2) . ومنطقة قبول هذا الفرض هي:

$$-Z_{\alpha/2} < Z < Z_{\alpha/2}$$

حيث α تساوي عادة 5% أو 1%.

الفرض الثاني: هو الفرض البديل H_1 والذي يعني وجود فرق معنوي بين متوسط العينة ومتوسط المجتمع أي أن $\mu - \mu_0$ حقيقي وغير ظاهري، وأن العينة تنتمي إلى مجتمع آخر له متوسط يختلف ولا يساوي متوسط المجتمع المعلوم، وعليه فإن:

$$H_1 : \mu \neq \mu_0$$

ومنطقة قبول هذا الفرض هي:

$$Z_{\alpha/2} \leq Z \leq -Z_{\alpha/2}$$

حيث α تساوي 5% أو 1%.

(2) تحديد درجة الثقة ومستوى المعنوية

إذا كانت درجة الثقة $1-\alpha = 95\%$ فإن مستوى المعنوية $\alpha = 5\%$ ، أما إذا كانت درجة الثقة $1-\alpha = 99\%$ فإن مستوى المعنوية $\alpha = 1\%$ وهكذا.

(3) تحديد التوزيع العيني لإحصائية الاختبار

ومنه نوجد Z - المحسوبة (*Z-calculated*) حيث:

$$Z_C = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

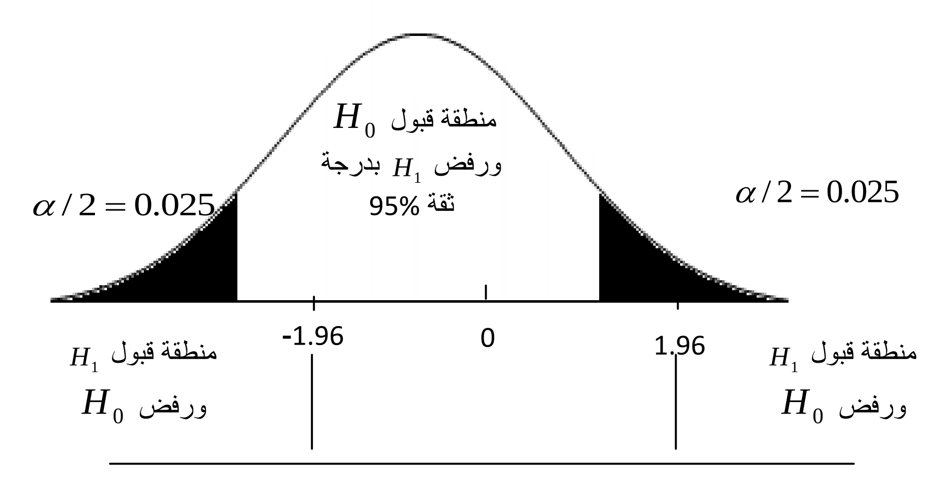
(4) إعطاء نتيجة الاختبار الإحصائي

إذا كانت مستوى المعنوية (α) تساوي 5% فإنه يمكن إيجاد القيم الحرجة التي على أساسها يتم الرفض والقبول. والقيم الحرجة يتم إيجادها كالتالي:

$$\pm Z_{\alpha/2} = \pm Z_{0.05/2} = \pm Z_{0.025} = \pm 1.96$$

وهنا إذا كانت قيمة Z - المحسوبة أصغر من 1.96 أو أكبر من 1.96- فإن قيمتها تقع في منطقة قبول الفرض الأول $H_0 : \mu = \mu_0$ بدرجة ثقة 95% ومستوى معنوية 5%، وهذا يعني أن العينة تمثل المجتمع وتنتمي إليه، وعليه فإننا نرفض الفرض الثاني $H_1 : \mu \neq \mu_0$. أما إذا كانت Z - المحسوبة أكبر من أو تساوي 1.96 أو أصغر من أو تساوي 1.96- فإن قيمتها تقع في منطقة قبول الفرض الثاني H_1 ومنطقة رفض الفرض الأول بنفس درجة الثقة ومستوى المعنوية وهذا معناه أن العينة لا تمثل المجتمع ولا تنتمي إليه. وهذا ما يوضحه الشكل التالي.

منطقة قبول H_0 ومنطقة رفض H_1 عند
مستوى معنوية 5%



شكل (رقم 17)

- **مثال (38)**: سحبت عينة عشوائية من 81 مفردة من مجتمع متوسطه 570وانحرافه المعياري 20، فوجد أن متوسط العينة هو 563. والمطلوب باحتمال 95% اختبار الفرض القائل أن هذه العينة مسحوبة من ذلك المجتمع ضد الفرض القائل أن متوسط هذه العينة لا يساوي متوسط المجتمع.

الحل:

(1)
$$H_0 : \mu = 570$$
$$H_1 : \mu \neq 570$$

(2) عند احتمال 95% وحيث أن الاختبار من طرفين، فإنه يمكن استخدام جداول التوزيع المعتدل العياري في إيجاد القيم الحرجة (انظر الشكل السابق) وهي تساوي ± 1.96.

(3) من العينة نجد أن:

$$Z_C = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{563 - 570}{\frac{20}{\sqrt{81}}} = -3.15$$

(4) حيث أن القيمة المحسوبة من العينة وقعت في منطقة الرفض فإننا نرفض الفرض العدم ونقبل الفرض البديل.

- **مثال (39)**: في المثال السابق نفرض أن $\bar{X} = 574$ وأن حجم العينة هو 121. والمطلوب باحتمال 99% اختبار فرض العدم القائل بأن متوسط المجتمع أقل من أو يساوي 570.

الحل:

(1)
$$H_0 : \mu \leq 570$$
$$H_1 : \mu > 570$$

(2) حيث أن الاختبار من طرف واحد فإنه باستخدام جداول التوزيع المعتدل العياري يمكن أن نوجد القيمة الحرجة كالتالي:

$$Z_\alpha = Z_{0.01} = 2.33$$

وهي القيمة التي يكون الاحتمال بعدها في الذيل الأيمن للتوزيع يساوي 1%.

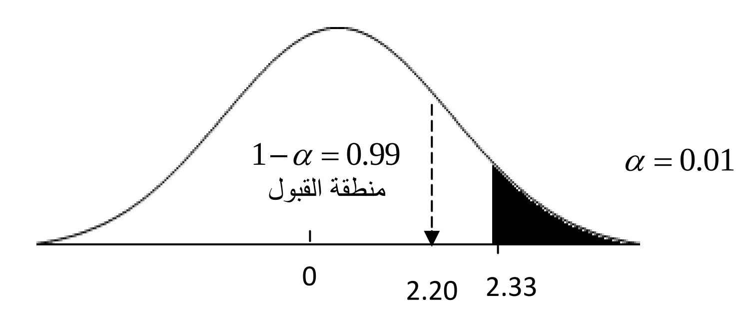
(3) باستخدام معلومات العينة نجد أن:

$$Z_C = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{574 - 570}{\frac{20}{\sqrt{121}}} = 2.20$$

(4) يقبل الفرض العدم القائل بأن هذه العينة مسحوبة من مجتمع متوسطه أقل من أو يساوي 570، وذلك لأن القيمة المحسوبة لـ Z وقعت في منطقة القبول.

منطقة قبول H_0 ورفض H_1 عند

مستوى معنوية 1%



شكل (رقم 18)

ثانيا: إذا كان تباين المجتمع غير معلوم

إذا كان تباين المجتمع غير معلوم عند إجراء اختبار المعنوية فإنه يمكن تقديره من العينة وذلك بحساب S^2، حيث:

$$S^2 = \frac{1}{n-1}\sum(X - \overline{X})^2$$

ولقد تم إثبات أن المتغير t :

$$t = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$$

يخضع لتوزيع t بدرجات حرية $n\text{-}1$ إذا كان حجم العينة n صغيرا (أي أقل من 30). أما إذا كان حجم العينة كبيرا (أي أكبر من 30) فإن المتغير t يتبع التوزيع المعتدل المعياري وإن كان تباين المجتمع σ^2 مجهولا.

● **مثال (40):** نفرض أن متوسط أعمار مصابيح شركة معينة هو 600 ساعة، ونفرض أن عينة أخذت بحجم 25 مصباحا وحسب متوسط الأعمار فيها فوجد أنه 630 ساعة والانحراف المعياري 37 ساعة. المطلوب باحتمال 95% اختبار ما إذا كانت هذه المصابيح مأخوذة من مجتمع متوسطه أكبر من أو يساوي 650.

الحل:

(1)
$$H_0 : \mu \geq 650$$
$$H_1 : \mu < 650$$

(2) حيث أن الانحراف المعياري للمجتمع غير معلوم وأن حجم العينة صغير وحيث أن الاختبار من طرف واحد وهو الطرف الأيسر، فإنه يمكن استخراج القيمة الحرجة من جدول t كما يلي: $-t_{\alpha,n-1} = -t_{0.05,24} = -1.71$.

(3) باستخدام العينة تحسب t - المحسوبة كالآتي:

$$t_C = \frac{630 - 650}{\frac{37}{\sqrt{25}}} = -2.70$$

(4) يرفض الفرض العدم ويقبل البديل القائل بأن متوسط المجتمع المحسوبة منه هذه العينة هو أقل من 650 ساعة طالما أن قيمة t - المحسوبة وقعت في منطقة الرفض.

- **مثال (41)**: أراد أحد الباحثين الاجتماعيين دراسة مستوى دخول مجموعة من الأسر بمدينة معينة، فقام بسحب عينة من 64 أسرة عشوائيا من هذه المجموعة فوجد أن الدخل في الأسرة المختارة يتوزع بمتوسط شهري قدره 43.25 دولارا وبانحراف معياري قدره 5.9 دولارا. فهل تعتقد أن مستوى دخل هذه المجموعة يتفق مع مستوى الدخل العام في تلك المدينة والذي يساوي 45 دولارا شهريا بدرجة ثقة 99%.

الحل:

(1)

$$H_0 : \mu = 45$$
$$H_1 : \mu \neq 45$$

(2) حيث أن الاختبار من طرفين وان التباين في المجتمع غير معلوم وأن حجم العينة كبير فانه يمكن استخدام جداول التوزيع الطبيعي المعياري لإيجاد القيمة الحرجة وهي:

$$\pm Z_{\alpha/2} = \pm Z_{0.01/2} = \pm Z_{0.005} = \pm 2.57$$

(3) بحساب قيمة Z من العينة نجد الآتي:

$$Z_c = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} = \frac{(43.25 - 45)\sqrt{64}}{5.9} = -2.37$$

(4) ونظرا لأن $Z_c > -Z_{0.005}$، فإنها تقع في منطقة القبول للفرض H_0 أي في منطقة رفض H_1. وتأسيسا على ذلك نقول أن مستوى دخول هذه المجموعة لا يختلف اختلافا حقيقياً عن مستوى دخول أسر تلك المدينة بدرجة ثقة 99% ومستوى معنوية 1%.

(2.2.4) اختبار النسبة في العينة (*Test for a Sample Proportion*)

إذا كنا بصدد مجتمع ينقسم إلى قسمين، قسم يمتلك صفة معينة وآخر لا يمتلك هذه الصفة مثل مجتمع وحدات الإنتاج التي تنقسم إلى وحدات معيبة وأخرى غير معيبة، فان النسبة في المجتمع سوف تكون مجهولة وباستخدام عينة عشوائية مسحوبة من ذلك المجتمع تحسب النسبة في العينة وتستخدم المعلومات المتاحة من العينة لاختبار بعض الفروض التي تتعلق بالنسبة في المجتمع. وفي هذا النوع من الاختبار يجب أن نفرق بين حالتين:

<u>الحالة الأولى</u>: عندما يكون حجم العينة صغيرا ($n<25$) وعندها يمكن أن نستخدم توزيع ذي الحدين في عمليات الرفض والقبول. فإذا أعطينا الرمز p لنسبة امتلاك الصفة في المجتمع حيث $p = 0.5$، فان X التي تمثل عدد الذين يملكون هذه الصفة في العينة التي حجمها n يتبع توزيع ذي الحدين إذا كان السحب بإرجاع بالخواص التالية:

$$E(X) = np_0$$
$$Var(X) = np_0q_0$$

وأن المتغير Z :

$$Z = \frac{X - 0.5 - np_0}{\sqrt{np_0q_0}}$$

يتبع التوزيع المعتدل المعياري إذا كان فرض العدم صحيحا.

<u>الحالة الثانية</u>: عندما يكون حجم العينة أكبر من 25 وكانت $p \neq 0.5$، حينها يمكن استخدام التوزيع المعتدل المعياري في عمليات الرفض والقبول حيث تم إثبات أن المتغير Z

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

يخضع للتوزيع المعتدل المعياري، وان $\hat{p}$ تمثل نسبة امتلاك الصفة في العينة التي حجمها n وان p_0 هي قيمة معينة معلومة.

- **مثال (42)**: سحبت عينة تتكون من 20 شخصا فوجد أن عدد المدخنين فيها 14 مدخنا. المطلوب اختبار الفرض القائل بان نسبة المدخنين في المجتمع أقل من أو يساوي 0.5 باحتمال 95%.

الحل:

(1)

$$H_0 : p \le 0.5$$
$$H_1 : p > 0.5$$

(2) حيث أن $\alpha = 5\%$ والاختبار من طرف واحد فان القيمة الحرجة التي تحدد مناطق الرفض والقبول هي 1.64 وذلك باستخدام جداول التوزيع المعتدل المعياري.

(3) حيث أن :

$$Z_c = \frac{X - 0.5 - np_0}{\sqrt{np_0 q_0}}$$

تتبع توزيعا معتدلا معياريا، فإننا سوف نقبل فرض العدم عندما تكون $Z<1.64$ ونقبل الفرض البديل عندما تكون $Z \ge 1.64$. وباستخدام بيانات العينة نجد أن:

$$Z_c = \frac{14 - 0.5 - (20)(0.5)}{\sqrt{20(0.5)(1 - 0.5)}} = 1.56$$

(4) لأن $Z_c < Z_\alpha$ ، فإننا نقبل فرض العدم والذي يعني أن هذه العينة مسحوبة من المجتمع الذي تكون فيه نسبة المدخنين أقل من أو يساوي 50% .

- **مثال (43)**: في شركة لبيع الملابس الجاهزة وجد أن 14% من الملابس المباعة يتم إرجاعها إلى الشركة بحجة أن مقاساتها غير مضبوطة. وفي محاولة للتغلب على

هذه المشكلة قامت الشركة بتغيير اسلوبها لتعديل المقاسات وأخذت عينة من 400 مشتر لهذا النوع من الملابس فوجدت أن 44 شخصا طلبوا تغيير الملابس المشتراة لعدم ملائمة المقاسات. المطلوب باحتمال 95% معرفة ما إذا كان هناك تحسين قد حدث في أداء الشركة من حيث ملائمة المقاسات.

الحل:

(1)
$$H_0 : p = 0.14$$
$$H_1 : p < 0.14$$

(2) عند مستوى 5% خطأ من النوع الأول، فانه باستخدام التوزيع المعتدل المعياري نجد أن القيمة الحرجة هي 1.64 − .

(3) حيث أن

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

فانه باستخدام المعلومات من العينة نجد أن $\hat{p} = 44/400 = 0.11$ وأن قيمة Z − المحسوبة هي:

$$Z = \frac{0.11 - 0.14}{\sqrt{\frac{0.14(1-0.14)}{400}}} = -1.76$$

(4) وحيث أن $Z_c < -1.64$ ، يرفض الفرض العدم ويقبل الفرض البديل القائل بان متوسط المجتمع المسحوب منه العينة أقل من 0.14 ، ألأمر الذي يؤشر على حدوث تحسن في مواصفات الإنتاج.

(3.2.4) اختبار التباين في العينة (*Test for a Sample Variance*)

أفرض أن لدينا مجتمعا معينا سحبت منه عينة من الحجم n وتم حساب التباين فيها ووجد أنه S^2 . المطلوب هو اختبار هل هذه العينة مسحوبة من مجتمع تباينه σ^2 أم لا ؟ أي اختبار الفرض $H_0 : \sigma^2 = \sigma_0^2$ حيث σ_0^2 قيمة محددة معلومة. ولإجراء هذا الاختبار فقد تم إثبات أن المتغير:

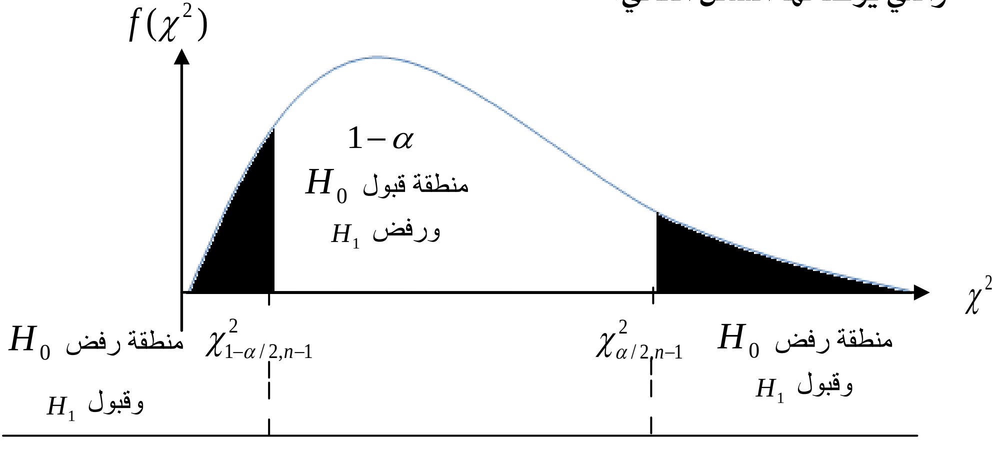
$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

يتبع توزيع χ^2 بدرجات حرية $n-1$. وباستخدام احتمال معين وجداول χ^2 فإنه يمكن تحديد القيم الحرجة وبالتالي مناطق الرفض والقبول. فإذا استخدمنا مستوى المعنوية $\alpha = 5\%$ فإن منطقة الرفض ستكون:

(أ) إذا كان البديل هو $H_1 : \sigma^2 \neq \sigma_0^2$

حيث σ_0^2 هي قيمة محددة، فنحن بصدد الاختبار من طرفين وأن منطقة الرفض سوف تكون: $\chi^2_{\alpha/2,n-1} \leq \chi^2 \leq \chi^2_{1-\alpha/2,n-1}$

والتي يوضحها الشكل التالي.



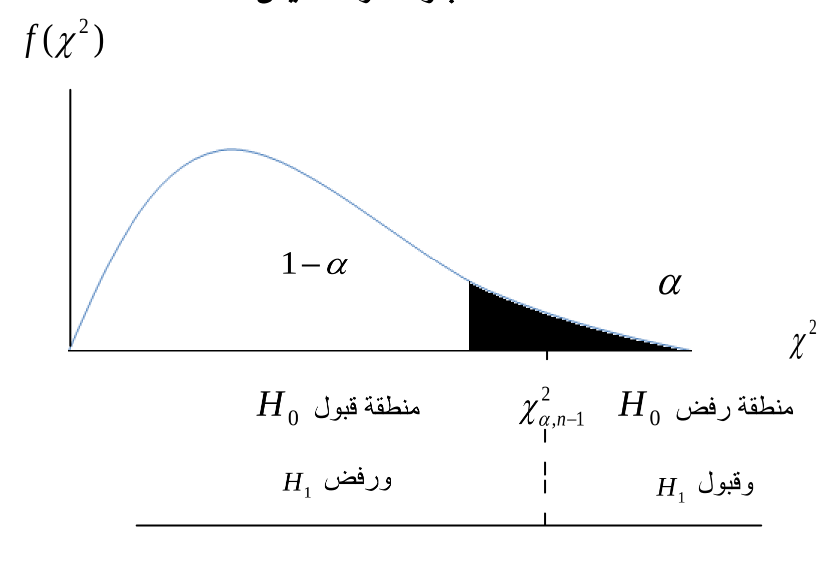
شكل (رقم 19)

(ب) إذا كان الفرض البديل هو

$$H_1 : \sigma^2 > \sigma_0^2$$

فهذا هو اختبار الطرف الأيمن، وإن منطقة الرفض سوف تكون: $\chi^2 \geq \chi^2_{\alpha,n-1}$ التي يوضحها الشكل التالي.

اختبار الطرف الأيمن



شكل (رقم 20)

(ج) أما إذا كان الفرض البديل هو

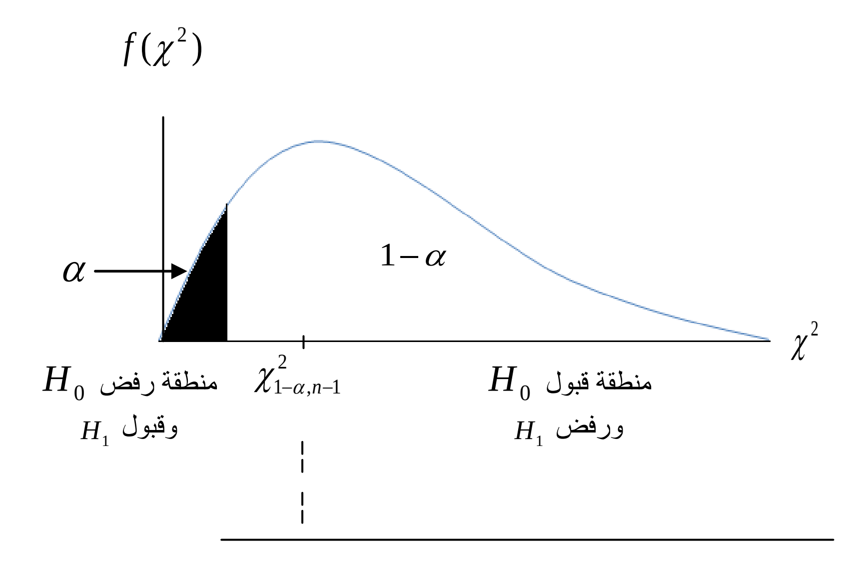
$$H_1 : \sigma^2 < \sigma_0^2$$

حيث نكون بصدد الطرف الأيسر، فإن منطقة الرفض هنا ستكون:

$$\chi^2 \leq \chi^2_{1-\alpha,n-1}$$

كما في الشكل التالي.

اختبار الطرف الأيسر



شكل (رقم 21)

- **مثال (44):** من مجتمع معين تم سحب عينة من 10 وحدات، حسب انحرافها المعياري فكان (0.35). المطلوب اختبار هل هذه العينة مسحوبة من مجتمع تباينه 0.09 وذلك باحتمال 95%.

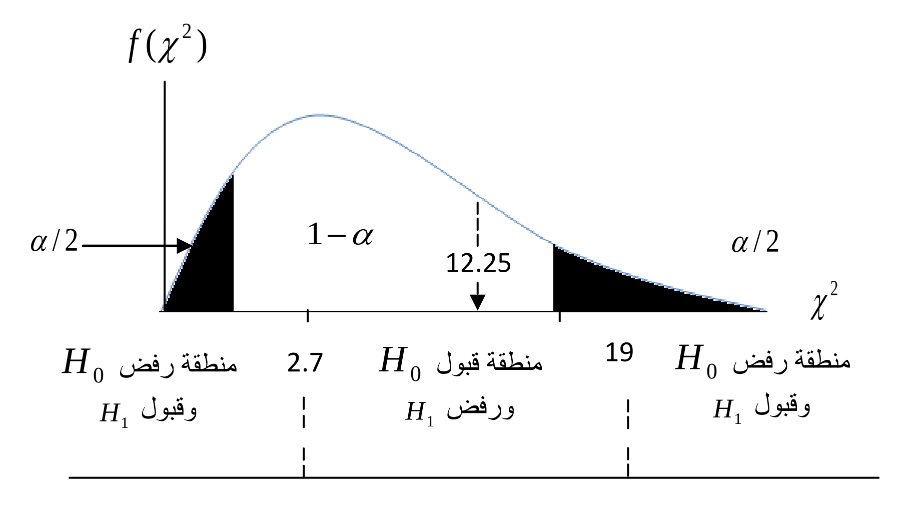
الحل:

(1)
$$H_0 : \sigma^2 = 0.09$$
$$H_1 : \sigma^2 \neq 0.09$$

(2) حيث أن الاختبار من طرفين فإنه باستخدام جداول χ^2 نجد أن القيم الحرجة بالشكل التالي:

$$\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,9} = 2.7$$
$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,9} = 19.0$$

فإذا وقعت قيمة المتغير χ^2 ما بين هذه الحدود فسوف نقبل الفرض العدم، بينما يقبل الفرض البديل إذا وقع المتغير χ^2 خارج هذه الحدود ويمكن تمثيلها في الشكل التالي.



شكل (رقم 22)

(3) باستخدام المعلومات من العينة نجد أن:

$$\chi_c^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(10-1)(0.35)^2}{0.09} = 12.25$$

(4) يقبل الفرض العدم القائل بأن هذه العينة مسحوبة من مجتمع تباينه 0.09، لأن قيمة $\chi^2 -$ المحسوبة تنحصر ما بين الحدين السابقين.

- **مثال (45)**: في المثال السابق اختبر الفروض الآتية:

$$H_0 : \sigma^2 = 0.09$$
$$H_1 : \sigma^2 > 0.09$$

الحل:

لأننا بصدد اختبار الطرف الأيمن، فإن قيمة $\chi^2 -$ المحسوبة تقارن مع قيمة $\chi^2 -$ الجدولية الآتية: $\chi^2_{\alpha,n-1} = \chi^2_{0.05,9} = 16.96$

ونظرا لأن قيمة $\chi_c^2 = 12.25$ ، وهي أصغر من قيمة $\chi^2 -$ الجدولية، فلا داعي إذن من رفض الفرض العدم وقبول الفرض البديل.

(4.2.4) اختبار معامل الارتباط في العينة

Test for sample coefficient of correlation

أفرض أن لدينا مجتمعا ذو بعدين *X, Y* يخضع للتوزيع الطبيعي ومعامل ارتباطه ρ ، أخذت منه المجموعة المكونة من *n* من الأزواج المرتبة وحسب معامل الارتباط لها فوجد أنه *r* وهو تقدير للمعامل ρ ، فإنه يمكننا أن نختبر ما إذا كانت هذه العينة مسحوبة من مجتمع له معامل ارتباط معين أم لا، بعد أن نفرق بين الحالتين التاليتين:

<u>الحالة الأولى:</u> إذا كان فرض العدم ينص على أن معامل الارتباط في المجتمع ρ يساوي صفرا. وهنا يكون فرض العدم بالشكل التالي:

$$H_0 : \rho = 0$$

مقابل واحد من الفروض البديلة الآتية:

$$H_1 : \rho \neq 0$$
$$H_1 : \rho > 0$$
$$H_1 : \rho < 0$$

وقد تم إثبات أن المتغير *t* حيث:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

يتبع توزيع *t* بدرجات حرية *n-2*، حيث *r* هو معامل الارتباط في العينة.

<u>الحالة الثانية:</u> إذا كان فرض العدم ينص على أن معامل الارتباط في المجتمع ρ له قيمة تختلف عن الصفر. وفي هذه الحالة فقد تم إثبات أن المتغير:

$$Z = (Z_1 - Z_2)\sqrt{n-3}$$

يتبع التوزيع المعتدل المعياري، حيث أن:

$$Z_1 = \frac{1}{2}\ln\frac{1+r}{1-r}, \qquad Z_2 = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}$$

- **مثال (46)**: سحبت عينة حجمها 65، حسب معامل الارتباط فيها فوجد انه يساوي 0.49 والمطلوب اختبار ما إذا كانت هذه العينة مسحوبة من مجتمع معامل الارتباط فيه يساوي صفرا وذلك باحتمال 95%.

الحل:

(1) نختبر الفرضية: $H_0 : \rho = 0$ مقابل الفرضية: $H_1 : \rho > 0$.

(2) باستخدام جدول التوزيع الطبيعي المعياري (حيث *n>30*) يمكن إيجاد القيمة الحرجة التي تحدد مناطق الرفض والقبول وهي تساوي 1.64 حيث $\alpha = 0.05$.

(3) من العينة تحسب قيمة *t* كما يلي :

$$t_c = \frac{(0.49)\sqrt{65-2}}{\sqrt{1-(0.49)^2}}$$

(4) نرفض فرض العدم ونقبل البديل القاضي بأن معامل الارتباط في المجتمع أكبر من صفر.

- **مثال (47)**: من عينة حجمها 50 تم حساب معامل الارتباط ووجد أنه يساوي 0.72، المطلوب اختبار هل هذه العينة مسحوبة من مجتمع معامل ارتباطه يساوي 0.85 وذلك باحتمال 99%.

الحل:

(1) نختبر الفرضية: $H_0 : \rho = 0.85$ مقابل الفرضية: $H_1 : \rho \neq 0.85$.

(2) باستخدام جدول المنحنى المعتدل المعياري وحيث أن الاختبار من طرفين، نوجد القيم الحرجة كما يلي:

$$\pm Z_{\alpha/2} = \pm Z_{0.01/2} = \pm Z_{0.005} = \pm 2.58$$

(3) تحسب قيمة Z من العينة كالتالي:

$$Z_1 = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1+0.70}{1-0.70} = 0.908$$
$$Z_2 = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.85}{1-0.85} = 1.256$$
$$Z_c = (Z_1 - Z_2)\sqrt{n-3}$$
$$= (0.908 - 1.256)\sqrt{50-3} = -2.39$$

(4) بما أن قيمة Z - المحسوبة أكبر من قيمة Z - الجدولية أي أن: -2.39>- 2.58 ، فان هذا يعني قبول الفرض العدم أو القبول بأن $\rho = 0.85$.

(5.2.4) اختبار معامل انحدار عينة

Test for Sample Coefficient of Regression

أفرض أن لدينا الخط المقدر لانحدار y على x الآتي:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

حيث $\hat{\beta}$ هو تقدير قيمة معلمة المجتمع المجهولة β باستخدام طريقة المربعات الصغرى،فان المطلوب اختبار هل أن معامل الانحدار في المجتمع المسحوبة منه العينة يساوي مقدارا معينا أم لا، وذلك باستخدام اختبارات الفروض. وفي هذا المجال فقد تم إثبات أن المتغير t

$$t = \frac{\hat{\beta} - \beta_0}{S(\hat{\beta})}$$

يتبع توزيع t بدرجات حرية n-2 ، حيث أن β_0 هي قيمة معينة منصوص عليها في فرض العدم وان $S(\hat{\beta})$ هي قيمة الانحراف المعياري لمعامل انحدار العينة $\hat{\beta}$ والذي تعرفه المعادلة التالية:

$$S(\hat{\beta})=\sqrt{\frac{S_{y/x}^{2}}{\sum(x-\bar{x})^{2}}}=\sqrt{\frac{MSE}{\sum x^{2}-n\bar{x}^{2}}}$$

حيث MSE هو متوسط مربعات الأخطاء (*Mean Square Error*). وبصورة أسهل فان :

$$S(\hat{\beta})=\sqrt{\frac{S_{yy}-\hat{\beta}^{2}S_{xx}}{(n-2)S_{xx}}}$$

حيث

$$S_{yy}=\sum(y-\bar{y})^{2}=\sum y^{2}-n\bar{y}^{2}$$
$$S_{xx}=\sum(x-\bar{x})^{2}=\sum x^{2}-n\bar{x}^{2}$$

- **مثال (48)**: من مجتمع معين سحبت عينة حجمها 5 فكانت قيم الأزواج المتناظرة لكحيل من (x_i, y_i) كما يلي:

x	3	4	5	2	1
y	5	8	7	6	4

والمطلوب اختبار هل هذه العين مسحوبة من مجتمع معامل الانحدار فيه يساوي صفرا أم لا وذلك باحتمال 95%.

الحل: الجدول التالي يبين الحسابات اللازمة للحل.

جدول (رقم 22)

x	y	$x\,y$	x^2	y^2
3	5	15	9	25
4	8	32	16	64
5	7	35	25	49
2	6	12	4	36
1	4	4	1	16
$\sum=15$	**30**	**98**	**55**	**190**

والآن

$$\bar{x}=\frac{\sum x}{n}=\frac{15}{5}=3$$

$$\bar{y}=\frac{\sum y}{n}=\frac{30}{5}=6$$

$$S_{xx}=55-5(3)^2=10$$

$$S_{yy}=190-5(6)^2=10$$

وباستخدام طريقة المربعات الصغرى نجد أن معامل الانحدار في العينة هو:

$$\hat{\beta}=\frac{\sum xy-n\bar{x}\bar{y}}{\sum x^2-n\bar{x}^2}=\frac{98-5(3)(6)}{55-5(3)^2}=0.8$$

وان الانحراف المعياري لمعامل الانحدار $\hat{\beta}$ يمكن حسابه كما يلي:

$$S(\hat{\beta})=\sqrt{\frac{10-(0.8)^2(10)}{3(10)}}=0.35$$

ثم نقوم بإجراء الاختبار كالتالي:

(1)

$$H_0:\beta=0$$
$$H_1:\beta\neq 0$$

(2) باستخدام جداول t نوجد القيم الحرجة كما يلي:

$$\pm t_{\alpha/2,n-2}=\pm t_{0.025,3}=\pm 3.18$$

(3) تحسب قيمة t من العينة كالآتي:

$$t_c=\frac{\hat{\beta}-\beta_0}{S(\hat{\beta})}=\frac{0.8-0}{0.35}=2.29$$

(4) نقبل فرض العدم القائل أن معامل الانحدار في المجتمع يساوي صفرا.

- **مثال (49)**: في المثال السابق، اختبر الفرض القائل أن $\beta = 1$ وذلك باحتمال 99%

الحل:

(1)
$$H_0 : \beta = 1$$
$$H_1 : \beta \neq 1$$

(2) القيم الحرجة هي: $\pm t_{\alpha/2,n-2} = \pm t_{0.005,3} = \pm 5.84$

(3) من العينة نجد أن:

$$t_c = \frac{\hat{\beta} - \beta_0}{S(\hat{\beta})} = \frac{0.8 - 1}{0.35} = -0.57$$

(4) نقبل الفرض العدم، أي قبول الفرضية $\beta = 1$ لأن t – المحسوبة تقع بين القيمتين الحرجتين ± 5.84 .

(3.4) اختبارات تتعلق بعينتين (*Tests Concerning Two Samples*)

وهي واحدة من أهم الاختبارات الإحصائية والتي عن طريقها نستطيع الحكم على وجود أو عدم وجود فرق معنوي بين معلمتي المجتمعين المسحوبة منهما العينتان. فإذا كان لدينا مجتمعان الأول يتوزع توزيعا معتدلا $N(\mu_1, \sigma_1^2)$ ، سحبت منه عينة عشوائية وحسب متوسطها فكان $\bar{X}$ ، والثاني يتوزع هو الآخر توزيعا معتدلا $N(\mu_2, \sigma_2^2)$ ، سحبت منه عينة عشوائية وحسب متوسطها فكان $\bar{Y}$ ، وما نريد أن نعرفه في هذه الحالة هو: هل أن الفرق بين متوسطي المجتمعين هو فرق معنوي (أي حقيقي وجوهري) أم أن الفرق بينهما راجع إلى مجرد الصدفة وناتج عن أخطاء المعاينة.

افرض أن باحثا اجتماعيا أراد على سبيل المثال معرفة إلى أي مدى يؤثر استخدام أسلوب جديد للتربية النفسية في رفع المستوى التعليمي للطلاب. في هذه الحالة يقوم الباحث بسحب عينة من طلاب المدارس ويستخدم معهم أسلوب التربية القديم ويسجل ملاحظاته على شكل درجات لكل طالب، ثم يختار مجموعة أخرى من طلاب المدرسة

ومن نفس الفئة العمرية ويستخدم معهم أسلوب التربية الحديث ويسجل مرة أخرى ملاحظاته على شكل درجات. وهنا تكون المشكلة هو اختبار ما إذا كان الفرق بين نتائج الأسلوب الحديث والأسلوب القديم فرقا معنويا، أي هل أن الفرق ناتج عن استخدام الأسلوب الحديث أم أن الفرق ظاهري وراجع إلى مجرد الصدفة وان الأسلوب الحديث لم يؤد إلى إحداث فرق جوهري وحقيقي بين نتائج الأسلوبين. ولا يقتصر الأمر فقط على الفرق بين المتوسطات فهناك اختبارات للنسبة واختبارات للتباين ولمعامل الارتباط.. الخ.

(1.3.4) اختبار الفرق بين نسبتي عينتين

أفرض أن لدينا عينتين كبيرتين ألأولى حجمها n وان نسبة الذين يملكون خاصية معينة فيها $\hat{p}_1$ ، والثانية حجمها m وان نسبة الذين يملكون نفس الخاصية السابقة فيها $\hat{p}_2$. أفرض أن نسبة الذين يملكون نفس الخواص في المجتمعين المسحوبة منه العينتان هي p_1 , p_2 على التوالي، فان المطلوب باستخدام هذه المعلومات اختبار هل هناك فرقا معنويا بين العينتين أم لا.

وتشمل فرضية العدم في هذه الحالة على أن الفرق بين نسبتي المجتمعين المسحوبة منهما العينتان تساوي قيمة معينة، أي أن

$$H_0 : p_1 - p_2 = d_0$$

وسوف نناقش بهذا الصدد الحالتين التاليتين:

الأولى: عندما تكون $d_0 = 0$: وهنا تصبح فرضية العدم على الشكل التالي:

$$H_0 : p_1 = p_2 = p$$

أما الفرض البديل فهو واحد من الأشكال التالية:

$$H_1 : p_1 \neq p_2$$
$$H_1 : p_1 > p_2$$
$$H_1 : p_1 < p_2$$

بحيث إذا تم قبول الفرض العدم فان هذا يعني أن الفرق الموجود بين العينتين فرق يعود إلى عملية المعاينة أو يعود إلى الصدفة والعكس صحيح في حالة قبولنا للفرض البديل.

ولإجراء الاختبار يتبقى أن نتعرف على التوزيع العيني الذي سيحكم هذه العملية والذي على أساسه سوف نحدد مناطق الرفض والقبول التي تعطي خطأ معينا من النوع الأول والثاني. ولقد تم إثبات أن المتغير Z

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\tilde{p}\tilde{q}(\frac{1}{n} + \frac{1}{m})}}$$

يتبع توزيعا معتدلا معياريا إذا كانت حجوم العينات $n+ m$ كبيرة ، حيث أن:

$$\hat{p}_1 = \frac{x}{n} \quad , \quad \hat{p}_2 = \frac{y}{m} \quad , \quad \tilde{p} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m} = \frac{x+y}{n+m}$$

- **مثال (50)**: عينتان، الأولى حجمها 100 مشاهدة وحجم الثانية 200 مشاهدة. أفرض أن عدد المدخنين في العينة الأولى هو 50 وفي العينة الثانية 120. المطلوب باحتمال 95% اختبار أن نسبة المدخنين في المجتمعين المسحوبة منهما العينتان متساوية.

الحل:

(1)
$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

(2) حيث أن الاختبار من طرفين والاحتمال هو 95%، فإنه وباستخدام جدول التوزيع الطبيعي المعياري نحدد مناطق الرفض والقبول كما يلي:

$$\pm Z_{\alpha/2} = \pm Z_{0.025} = \pm 1.96$$

(3) باستخدام المعلومات من العينتين نجد الآتي:

$$\hat{p}_1 = \frac{x}{n} = \frac{50}{100} = 0.5, \quad \hat{p}_2 = \frac{y}{m} = \frac{120}{200} = 0.6$$

حيث أن $\hat{p}_2, \hat{p}_1$ تقديرات غير متحيزة لكل من المعلمتين p_2, p_1 على التوالي.

$$\tilde{p} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m} = \frac{0.5(100) + 0.6(200)}{100 + 200} = 0.57$$

$$\therefore \tilde{q} = 1 - \tilde{p} = 1 - 0.57 = 0.43$$

وبالتالي فإن:

$$Z_C = \frac{0.5 - 0.6}{\sqrt{(0.57)(0.43)(\frac{1}{100} + \frac{1}{200})}} = -1.67$$

(4) وحيث أن $Z-$ المحسوبة تقع في منطقة القبول نقبل الفرض العدم ونرفض البديل ونستنتج أن نسبة المدخنين في المجتمعين متساوية.

- **مثال (51)**: في المثال السابق المطلوب اختبار الفروض الآتية:

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 < p_2$$

الحل:

لأن الاختبار من طرف واحد وأن $\alpha = 0.05$ ، فإن تحديد منطقة الرفض سوف تكون كما يلي:

$$-Z_\alpha = -Z_{0.05} = -1.65$$

القرار: يرفض H_0 لأن قيمة $Z_c = -1.67$ واقعة في منطقة الرفض، لذا فإن نسبة المدخنين في العينة الأولى سوف تكون أقل مما هي عليه في العينة الثانية.

<u>**الحالة الثانية:**</u> عندما تكون $d_0 \neq 0$: وهنا ستكون فرضية العدم كالتالي:

$$H_0 : p_1 - p_2 = d_0$$

أي أن p_1, p_2 مختلفتان. أما الفرض البديل فسوف يأخذ إحدى الحالات التالية:

$$H_1 : p_1 - p_2 \neq d_0$$
$$H_1 : p_1 - p_2 > d_0$$
$$H_1 : p_1 - p_2 < d_0$$

وقد وجد أن المتغير العشوائي Z

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - d_0}{\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}}$$

يتبع التوزيع الطبيعي المعياري عندما تكون H_0 صحيحة. وبما أن p_1, p_2 غير متساوية كما ينص الفرض العدم، فإننا نحسب التقديرات $\hat{p}_1$ و $\hat{p}_2$ وبذلك فإن قيمة Z سوف تكون:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - d_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}}}$$

- **مثال (52):** أدعى أحد المدراء بعدم وجود أي اختلاف في نسب المعيب بين خطين من خطوط الإنتاج بالمصنع تحت الدراسة. وللتأكد من ذلك قام أحد الإحصائيين بسحب عينة عشوائية من 35 وحدة من إنتاج الخطين الأول والثاني، فوجد أن نسبة المعيب في هذا الإنتاج هو 0.29 ,0.277 على التوالي. عند مستوى معنوية 5% أختبر أن نسبة المعيب في إنتاج الخط الأول يزيد معنويا عن نسبة المعيب في إنتاج الخط الثاني بـ 0.011.

الحل:

(1)

$$H_0 : p_1 - p_2 = 0.011$$
$$H_1 : p_1 - p_2 > 0.011$$

(2) نحن هنا بصدد اختبار الطرف الأيمن. وحيث أن $\alpha = 0.05$ فإنه باستخدام جداول المنحنى الطبيعي المعياري يمكن أن نحدد المنطقة الحرجة كالآتي:

$$Z_\alpha = Z_{0.05} = 1.65$$

(3) من العينة نجد أن:

$$\hat{p}_1 = 0.29 \quad , \quad \hat{p}_2 = 0.277$$

وبالتالي فإن قيمة المختبر الإحصائي Z هي:

$$Z_c = \frac{(0.29 - 0.277) - 0.011}{\sqrt{(\frac{(0.29)(0.71)}{35} + \frac{(0.277)(0.723)}{35}}} = 0.0185$$

(4) وحيث أن قيمة $Z-$ المحسوبة تقع في منطقة القبول لذلك يقبل الفرض العدم ويرفض البديل الذي يقضي بوجود فرق بخطوط الإنتاج في نسبة المعيب بكل واحد منهما. وإذا كان هناك من فرق فهو يعزى للعمل العشوائي فقط.

(2.3.4) اختبار نسبة تباينين

في مواقف كثيرة يكون اهتمامنا باختبار فروض تتعلق بتباين عينتين أكثر من أي معلمة أخرى. ولإجراء هذا الاختبار نفرض أن لدينا عينتين الأولى حجمها n وتباينها S_1^2 مسحوبة من مجتمع تباينه σ_1^2، والثانية حجمها m وتباينها S_2^2 مسحوبة من مجتمع تباينه σ_2^2، والمطلوب باستخدام هذه المعلومات اختبار ما إذا كان هناك فرق معنوي بين تباين المجتمعين أم لا؟ معنى ذلك أن فرض العدم سوف يكون كالتالي:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

على أن يكون فرض البديل واحدا من الفروض الآتية:

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$
$$H_1 : \sigma_1^2 < \sigma_2^2$$

وعندما نقبل الفرض العدم فهذا يعني أن العينتين مسحوبتان من مجتمع واحد، وفي هذه الحالة يمكن القول أن العينتين متجانستان (*Homogenous*). أما إذا رفضنا الفرض العدم وقبلنا الفرض البديل فإننا نستنتج أن العينتين مسحوبتان من مجتمعين مختلفين، وبذلك تكون العينتان (*Heterogeneous*). ولذلك يطلق على هذا الاختبار أحيانا اسم اختبار التجانس (*Test of homogeneity*) وهو اختبار ضروري جدا قبل إجراء بعض الاختبارات التي تتعلق بالمتوسطات كما سنرى لاحقا. مرة أخرى، إذا قبلنا الفرض العدم فان ذلك يعني أن التباين في

المجتمعين متساوي ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) وأنه يمكن تقدير قيمة مشتركة للتباين S_P^2 تسمى بالتباين المشترك وتعرف كما يلي:

$$S_P^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

حيث تعتبر S_P^2 تقدير جيد لكل من σ_2^2, σ_1^2 وبالتالي لـ σ^2. أما التوزيع العيني الذي عن طريقه سوف نحدد مناطق الرفض والقبول أو الذي على أساسه سوف نجري الاختبار فقد تم إثبات أن المتغير F حيث

$$F_c = \frac{S_1^2}{S_2^2}$$

يتبع توزيع F بدرجات حرية $(n\text{-}1)$, $(m\text{-}1)$ إذا كان $S_1^2 > S_2^2$. أما إذا كان $S_2^2 > S_1^2$ فقد وجد أن المتغير:

$$F_c = \frac{S_2^2}{S_1^2}$$

يتبع هو الآخر توزيع F ولكن بدرجات حرية $(m-1), (n-1)$. هذه النسبة تسمى بقيمة $F-$ المحسوبة. ومن جداول F نستخرج قيمة $f-$ الجدولية ثم نستنتج الآتي:

(1) إذا كان الفرض البديل هو $H_1:\sigma_1^2 \neq \sigma_2^2$ ، فهذا يعني أن الاختبار من طرفين وأن مناطق الرفض هي:

$$f_{\alpha/2,(n-1,m-1)} \leq F_c \leq f_{1-\alpha/2,(n-1,m-1)}$$

$$f_{\alpha/2,(n-1,m-1)} \leq F_c \leq \frac{1}{f_{\alpha/2,(m-1,n-1)}}$$

من ناحية ثانية فإنه يمكن اعتبار هذا الاختبار هو اختبار (طرف أيمن) مع أخذ *F* عند $\alpha/2$ بدلا من α (أي $f_{\alpha/2,(n-1,m-1)}$ بدلا من $f_{\alpha,(n-1,m-1)}$). وهنا إذا كانت قيمة *f*– الجدولية أكبر من قيمة *F*– المحسوبة يكون الفرق غير معنوي، ونقبل الفرض العدم ونسلم بأن هناك تجانسا وأنه يمكن تقدير تباين مشترك للمجتمعين.

(2) إذا كان الفرض البديل هو:

$$H_1:\sigma_1^2 > \sigma_2^2$$

فإن منطقة الرفض سوف تكون

$$F_c \geq f_{\alpha,(n-1,m-1)}$$

(3) وأخيرا إذا كان الفرض البديل هو:

$$H_1:\sigma_1^2 < \sigma_2^2$$

فإن منطقة الرفض سوف تكون كالآتي

$$F_c \leq \frac{1}{f_{\alpha,(m-1,n-1)}}$$

- **مثال (53):** عينتان الأولى حجمها *n=10* وتباينها $S_1^2 = 7.14$، والثانية حجمها *n=8* وتباينها $S_2^2 = 3.21$. والمطلوب باحتمال 95% اختبار التجانس بين المجموعتين، أي اختبار الفرض $H_0:\sigma_1^2 = \sigma_2^2$ مقابل الفرض $H_1:\sigma_1^2 \neq \sigma_2^2$.

الحل: يمكننا أن نحل المثال بطريقتين:

الطريقة الأولى: وفيها نعتبر أن الاختبار من طرف واحد وهو الطرف الأيمن. وحيث أن S_1^2 أكبر من S_2^2 فإننا نستخرج من جداول *F* القيمة الجدولية عند $\alpha/2$ (حيث $\alpha = 0.05$) بدرجات حرية $(n-1), (m-1)$ كما يلي:

$$F_{\alpha/2,(n-1,m-1)} = f_{0.025,(9,7)} = 4.82$$

ومن العينة تحسب *F*– المحسوبة كالآتي:

$$F_c = \frac{S_1^2}{S_2^2} = \frac{7.14}{3.21} = 2.22$$

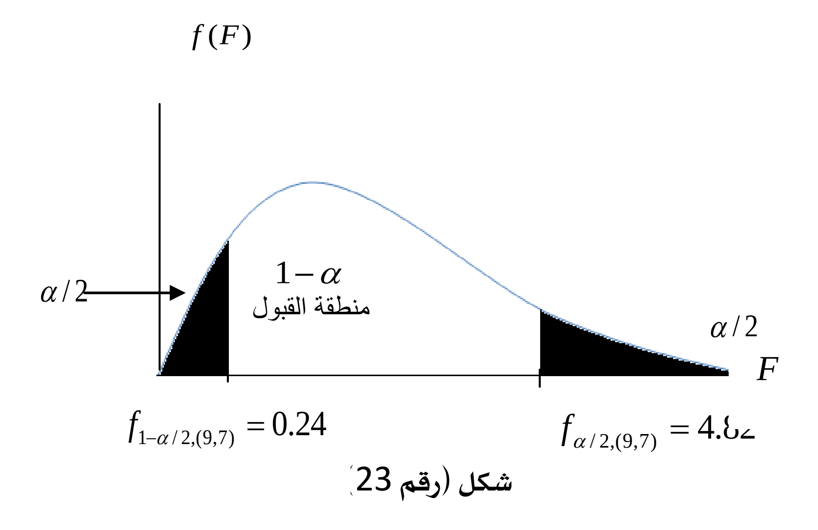
وحيث أن F_c أصغر من *f*– الجدولية يقبل الفرض العدم ويرفض البديل، وهذا يعني وجود تجانس بين العينتين وبالتالي عدم وجود أي مبرر للقول بأن التباينين مختلفان وبذلك يمكننا تقدير تباين مشترك للمجتمعين إذا شئنا ذلك.

الطريقة الثانية: بموجب هذه الطريقة فإننا نوجد القيمتين الحرجتين الآتيتين:

$$f_{\alpha/2,(n-1,m-1)} = f_{0.025,(9,7)} = 4.82$$

$$f_{1-\alpha/2,(n-1,m-1)} = \frac{1}{f_{\alpha/2,(m-1,n-1)}} = \frac{1}{f_{0.025,(7,9)}} = \frac{1}{4.197} = 0.24$$

ونظرا لأن قيمة *F*– المحسوبة تقع بين هاتين القيمتين ولا تساوي أحدا منهما، فإنها تقع في منطقة القبول وعليه نقبل H_0 ونرفض H_1 وهذا ما يتضح لنا من الشكل التالي.



شكل (رقم 23)

●**مثال (54): في المثال السابق هل تعتقد أن تباين العينة الأولى أكبر معنويا من تباين العينة الثانية عند** $\alpha = 0.05$ **؟**

الحل: هنا نقوم بإجراء اختبار الطرف الأيمن لنسبة التباين $\frac{\sigma_1^2}{\sigma_2^2}$ **حيث نجد أن:**

(1)
$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$

(2) نوجد قيمة f– الجدولية كما يلي:

$$f_{\alpha,(n-1,m-1)} = f_{0.05,(9,7)} = 3.68$$

(3) من العينة وجدنا أن قيمة F- المحسوبة تساوي 2.22

(4) ونظرا لأن قيمة F – المحسوبة أصغر من f– الجدولية فإنها تقع في منطقة القبول، لذلك نقبل H_0 وأن المشاهدات تؤيد صحة الفرض بمعنى أن تباين العينة الأولى لا يزيد معنويا (من وجهة النظر الإحصائية) عن تباين العينة الثانية وذلك عند مستوى معنوية 5%.

(3.3.4) اختبار الفرق بين متوسطي عينتين

من المشاكل الشائعة الاستخدام في اختبارات الفروض هي الاختبارات الخاصة بمعرفة هل هناك فرق معنوي بين متوسطي عينتين أم لا؟ بمعنى آخر هل هاتان العينتان مسحوبتان من نفس المجتمع أم أنهما مسحوبتان من مجتمعين مختلفين؟ وفي هذه الحالة يأخذ الفرض العدم الشكل التالي:

$$H_0 : \mu_1 - \mu_2 = d_0$$

ويمكن أن يأخذ الفرض البديل واحدا من الأشكال التالية:

$$H_1 : \mu_1 - \mu_2 \neq d_0$$
$$H_1 : \mu_1 - \mu_2 > d_0$$
$$H_1 : \mu_1 - \mu_2 < d_0$$

حيث d_0 قيمة معينة معلومة. وفي حالة قبول الفرض العدم فإن هذا يعني أن العينتين مسحوبتان من نفس المجتمع وعند رفضه فلا يمكن القول أن العينتين مسحوبتان من نفس المجتمع أو أن لهما نفس المتوسط.

وعندما تكون $d_0 = 0$ فإن فرضية العدم السابقة سوف تكون كالتالي:

$$H_0 : \mu_1 = \mu_2$$

مقابل الفرضيات البديلة الآتية:

$$H_1 : \mu_1 \neq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

وسوف نتحدث عن هذه المشكلة تفصيلا بعد أن نفرق بين حالتين: حالة العينات المستقلة وحالة العينات الغير مستقلة. ويقال عن عينتين أنهما مستقلتان إذا كانت مفردات إحداهما مستقلة تماما عن مفردات العينة الثانية. أما إذا كانت هناك علاقة ما بين العينتين فإن هذا يعني عدم الاستقلال.

أولا: العينات الغير مستقلة (*Dependent Samples*)

العينات الغير مستقلة تكون في حالة عدم تغيير مفردات العينة. ونحن نقوم بإجراء الاختبار بعد إعطاء هذه المفردات معالجة من نوع جديد (من شأنها أن تقلل أو تزيد من القياسات التي تؤخذ لهذه المفردات) مثل الدواء الجديد الذي يقلل من ضغط الدم أو نسبة السكر أو إعطاء غذاء جديد يؤدي إلى زيادة الوزن أو طريقة تربوية جديدة تؤدي إلى دفع مستوى الطلاب أو سماد جديد يؤدي إلى زيادة إنتاجية الأرض الزراعية إلى غير ذلك.

وللتعرف على كيفية إجراء الاختبار في هذه الحالة نفرض أن لدينا عينتين غير مستقلتين حجم كل منهما n والمطلوب اختبار هل هناك فرق معنوي بين العينتين أم لا؟ نفرض أن الفرق بين القراءات في العينة الأولى والثانية هو d_i حيث أن:

$$d_i = X_i - Y_i, \qquad i = 1, 2, \ldots, n$$

علما بأن مفردات العينة الأولى هي $x_1, \ldots, x_n$ وأن $y_1, \ldots, y_n$ هي مفردات العينة الثانية. وفي هذه الحالة يأخذ الفرض العدم الشكل التالي:

$$H_0 : \mu_1 - \mu_2 = \mu_d$$

مقابل الفرض البديل الآتي:

$$H_1 : \mu_1 - \mu_2 \neq \mu_d$$

ولأن الفرق بين متوسط المجتمعين يكون مساويا لمتوسط الفروق، فإن فرض العدم يصبح كالتالي:

$$H_0 : \mu_d = d_0$$

مقابل الفرض البديل

$$H_1 : \mu_d \neq d_0$$

حيث d_0 كمية معينة معلومة. وبالتالي فإن الاختبار قد اختزل إلى اختبار عينة واحدة. وفي حالة

قبولنا لفرض العدم فهذا يعني قبولنا بعدم وجود فرق بين المجموعتين، أي ليس هناك فرق بين قيمة الظاهرة قبل وبعد تعاطي الدواء مثلا والعكس صحيح.

يبقى هنا أن نتعرف على التوزيع العيني الذي يحكم هذا الاختبار، وقد وجد أن المتغير t:

$$t = \frac{(\bar{d} - \mu_d)\sqrt{n}}{S_d}$$

يخضع لتوزيع t بدرجات حرية $(n\text{-}1)$ إذا كان حجم العينة صغيرا، حيث $\bar{d}$ هو متوسط الفروق وأن S_d هو الانحراف المعياري للفرق. أما إذا كانت n كبيرة فإنه يمكن استخدام التوزيع المعتدل المعياري لذلك.

- **مثال (55)**: قام أحد الباحثين في مجال الطب باختيار عينة تحتوي على 5 أطفال وسجل أوزانهم فكانت كالتالي: (10 ,9 ,8 ,12 ,11) كغم. وبعد تناول الأطفال غذاء من نوع جديد لمدة شهر واحد سجلت أوزانهم فكانت القراءات كالآتي: (11.5, 12.2, 9, 8.8, 10) كغم. فإذا كان المجتمع المسحوبة منه العينة يتوزع توزيعا معتدلا بتباين مجهول، المطلوب معرفة ما إذا كان الغذاء الجديد قد ساعد على زيادة وزن الطفل معنويا عند مستوى معنوية 1%.

الحل: لإجراء الاختبار نجد فرق القراءتين كالآتي:

$$d_i = -0.5,\ 0.2,\ -1,\ 0.2,\ 0$$

وأن متوسط الفرق $\bar{d}$ هو

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-1.5}{5} = -0.3$$

أما الانحراف المعياري لهذا الفرق فهو

$$S_d = \sqrt{\frac{1}{n-1}\left(\sum d^2 - n\bar{d}^2\right)}$$

$$S_d = \sqrt{\frac{1}{5-1}\left(1.33 - 5(-0.3)^2\right)} = 0.469$$

وهنا نضع فروض الاختبار على النحو الآتي:

(1) $H_0 : \mu_d = 0$ (الغذاء لم يؤد إلى زيادة الوزن معنويا)

$H_1 : \mu_d \neq 0$ (الغذاء أدى إلى زيادة الوزن معنويا)

(2) باستخدام جدول t نحدد مناطق الرفض والقبول وهي:

$$\pm t_{\alpha/2,n-1} = \pm t_{0.005,4} = \pm 4.60$$

(3) من العينة تحسب قيمة t كالتالي:

$$t_C = \frac{-0.3\sqrt{5}}{0.469} = -1.43$$

(4) وحيث أن قيمة t – المحسوبة تقع في منطقة القبول فإننا نقبل الفرض العدم ونرفض الفرض البديل، وهذا يعني عدم وجود اختلاف حقيقي بين متوسط الوزن قبل تناول الغذاء الجديد ومتوسط الوزن بعد تناول الغذاء الجديد.

ثانيا: العينات المستقلة (*Independent Samples*)

في هذه الحالة نفرض أن العينتين المراد اختبار الفرق بينهما عينتان مستقلتان أي أن مفردات العينة الأولى تختلف تماما عن مفردات العينة الثانية. وهنا يتوقف شكل التوزيع العيني وشكل الاختبار على: هل أن تباين المجتمع معلوم لكل من العينتين أم لا ؟ وعليه سوف نناقش كلا الحالتين فيما يأتي.

الحالة الأولى: حالة معلومية التباينات σ_2^2, σ_1^2

أفرض أن لدينا عينتين مستقلتين حجم الأولى n والثانية m مسحوبتان من مجتمعين متوسط الأول μ_1 ومتوسط الثاني μ_2 . فإذا كان متوسط العينة الأولى $\bar{X}$ ومتوسط العينة الثانية $\bar{Y}$ وأن تباين

المجتمع الأول σ_1^2 والثاني σ_2^2 معلومان، فإن المطلوب اختبار ما إذا كان متوسط المجتمع المسحوبة منه العينة الأولى يساوي متوسط المجتمع المسحوبة منه العينة الثانية.

ولاختبار هذا الفرض فإن المطلوب تحديد التوزيع العيني الذي يتبع له المتغير والذي عن طريقه يمكن قبول أو رفض بعض الفروض. ولقد تم إثبات أن المتغير Z حيث

$$Z_c = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

يتبع التوزيع المعتدل المعياري. وإذا كنا قد افترضنا في الفرض العدم تساوي متوسط المجتمع الأول والثاني، فإن المتغير Z سوف يأخذ الشكل التالي:

$$Z_c = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

- **مثال (56):** من أحد المصانع سحب باحث عينة من 81 عاملا فوجد أن متوسط إنتاجية العامل فيه تساوي 75 وحدة يوميا، ثم سحب عينة أخرى من مصنع آخر بحجم 169 عاملا فوجد أن متوسط إنتاجية هذه العينة تساوي 70 وحدة يوميا، بينما كان الانحراف المعياري لإنتاجية العمال في المصنع الأول 5 وحدات وفي المصنع الثاني 4 وحدات. وقد أقر الباحث باستخدام هذه البيانات بوجود فرق حقيقي بين إنتاجية العمال في المصنعين. فهل تتفق مع الباحث في ذلك بدرجة ثقة 95%؟

الحل:

(1)

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$

(2) حيث أن الاختبار من طرفين فإنه باستخدام جداول المنحنى المعتدل المعياري نستطيع أن نحدد مناطق الرفض والقبول عند $\alpha = 0.05$ وهي: ± 1.96 .

(3) من بيانات العينة يمكن حساب $Z-$ المحسوبة كما يلي:

$$Z_C = \frac{(75-70)}{\sqrt{\frac{25}{81}+\frac{16}{169}}} = 7.87$$

(4) نظرا لأن $Z_C > Z_{0.025}$ ، فإننا نرفض الفرض العدم القائل أن العينتين مسحوبتان من نفس المجتمع، ونقبل الفرض البديل القائل بأن متوسط المجتمع المسحوبة منه العينة الأولى يختلف عن متوسط المجتمع المسحوبة منه العينة الثانية. وهذا يعني وجود فرق بين إنتاجية العمال في المصنعين بدرجة ثقة 95% وتكون الإنتاجية في المصنع الثاني أقل من الإنتاجية في المصنع الأول لأن $\bar{X} > \bar{Y}$ ، وعليه فنحن نتفق مع الباحث في استنتاجه.

الحالة الثانية: حالة عدم معلومية التباينات σ_2^2 , σ_1^2

إذا كانت تباينات المجتمعين غير معلومة فإن شكل التوزيع العيني يتوقف على ما إذا كان هناك تجانس بين العينتين أم لا ؟ فإن كان هناك تجانس نجد أن التوزيع العيني في هذه الحالة يختلف عن التوزيع العيني عند حالة عدم وجود التجانس بين العينتين. وسوف ندرس كلا الحالتين في الآتي:

(1) **عند وجود تجانس:** إذا كان هناك تجانس بين العينتين فهذا معناه أن $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ، ولا يمكن استنتاج ذلك إلا بعد إجراء اختبار التجانس وذلك باستخدام اختبار F (راجع بند اختبار نسبة تباينين 2.3.4). أما التوزيع العيني الذي يمكن استخدامه في هذا الاختبار فهو توزيع t إذا كانت أحجام العينات صغيرة ($n + m < 50$). ولقد تم إثبات أن المتغير t حيث:

$$t_c = \frac{\bar{X}-\bar{Y}}{\sqrt{S^2(\frac{1}{n}+\frac{1}{m})}}$$

يتبع توزيع t بدرجات حرية $(n + m - 2)$. أما إذا كانت أحجام العينات كبيرة $(n + m>50)$ فسوف نستخدم جداول المنحنى المعتدل المعياري.

- **مثال (57)**: استخدم أحد الباحثين نوعان من الهرمونات لمعرفة أثرها على مرض معين فكانت لدينا النتائج الآتية:

هرمون a	57	120	101	137	119	117	104	73	53	98	110
هرمون b	79	30	82	50	39	22	57	32	96	31	88

والمطلوب اختبار الفرض القائل بأن هرمون (a) له نفس تأثير هرمون (b) وذلك باحتمال 95%

الحل: حيث أن تباينات المجتمعين غير معلومة فإننا نجري اختبار التجانس أولا كما يلي:

(1)
$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

(2) نحسب S_1^2 , S_2^2 كالآتي:

$$S_1^2 = \frac{1}{n-1}(\sum X^2 - n(\bar{X})^2)$$
$$= \frac{1}{11-1}(115127 - 11(99)^2) = 731.6$$

حيث أن $\bar{X} = 99$.

$$S_2^2 = \frac{1}{m-1}(\sum Y^2 - m(\bar{Y})^2)$$
$$= \frac{1}{11-1}(40564 - 11(55)^2) = 728.9$$

حيث أن $\bar{Y} = 55$.

وبما أن $S_1^2 > S_2^2$ ، فإننا نستخدم جداول F لإيجاد قيمة $f-$ الجدولية كما يلي:

$$f_{\alpha/2,(n-1,m-1)} = f_{0.025,(10,10)} = 3.72$$

(3) ومن العينة يمكن حساب $F-$ المحسوبة كالتالي:

$$F_c = \frac{S_1^2}{S_2^2} = \frac{731.6}{728.9} = 1.004$$

(4) وحيث أن قيمة $F-$ المحسوبة أقل من قيمة $f-$ الجدولية، نقبل الفرض العدم القائل بوجود التجانس بين العينتين. وعلى ذلك يمكن الاستمرار وإجراء اختبار الفرق بين المتوسطين بالشكل التالي:

(1)
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

(2) حيث أن الاختبار من طرفين، نوجد القيم الحرجة من جداول t وهي:

$$\pm t_{\alpha/2,(n+m-2)} = \pm t_{0.025,(11+11-2)} = \pm t_{0.025,20} = \pm 2.09$$

(3) من العينة يمكن حساب التباين المشترك باستخدام الصيغة التالية:

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

$$S_p^2 = \frac{(11-1)(731.6) + (11-1)(728.9)}{11+11-2} = 730.25$$

وبالتالي فان قيمة $t-$ المحسوبة هي:

$$t_c = \frac{99-55}{\sqrt{730.25(\frac{1}{11}+\frac{1}{11})}} = 3.819$$

(4) لأن قيمة t – المحسوبة تقع في مناطق الرفض، عليه نرفض الفرض العدم ونقبل الفرض البديل القائل بأن العينتين مسحوبتان من مجتمعين مختلفين أو بمعنى آخر أن الهرمون ألأول والثاني مختلفان في التأثير على المرض.

(2) عند عدم وجود تجانس: إذا قبل الفرض البديل في اختبار التجانس فهذا يعني أن العينتين غير متجانستين وبالتالي فان $\sigma_1^2 \neq \sigma_2^2$، ألأمر الذي لا يمكن معه تقدير التباين المشترك. وفي حالة كهذه فقد تم إثبات أن التوزيع العيني الذي يحكم الاختبار هو توزيع قريب من توزيع t يطلق عليه اسم توزيع t' ، حيث

$$t' = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$$

يخضع لتوزيع t' والذي ليس له جداول خاصة به ولكن يمكن استخدام جداول t العادية في إيجاد قيمة t – الجدولية بعد إجراء تعديل في درجات الحرية والتي تحسب بالمعادلة التالية:

$$d.f = \frac{(\frac{S_1^2}{n} + \frac{S_2^2}{m})^2}{\frac{(S_1^2/n)^2}{n-1} + \frac{(S_2^2/m)^2}{m-1}}$$

- **مثال (58):** استخدم نوعان من الفيتامينات في تغذية مجموعتين من الأطفال، فكانت النتائج الآتية:

العينة الثانية	العينة الأولى	
21	25	المتوسط
0.7	3.5	التباين
12	8	N

والمطلوب اختبار ما إذا كان هذان النوعان من الفيتامينات لهما نفس التأثير باحتمال 95%.

ألحل: نجري أولا اختبار التجانس كما يلي:

(1)
$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

(2) حيث أن $S_1^2 > S_2^2$ فإننا نستخرج قيمة F من الجدول كالتالي:

$$f_{\alpha/2,(n-1,m-1)} = f_{0.025,(7,11)} = 3.76$$

(3) من العينة نحسب

قيمة F كما يلي:
$$F_c = \frac{S_1^2}{S_2^2} = \frac{3.5}{0.7} = 5$$

(4) حيث أن $f -$ الجدولية أقل من قيمة $F -$ المحسوبة، نقبل الفرض البديل القائل بأن المجموعتين غير متجانسين، ولذلك نجري اختبار t' كالتالي:

(1)
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

(2) لتحديد مناطق الرفض والقبول نستخرج قيمة t أولا من جدول t بعد أن نحدد درجات الحرية ($d.f$) كالتالي:

$$d.f = \frac{\left(\frac{3.5}{8} + \frac{0.7}{12}\right)^2}{\frac{(3.5/8)^2}{7} + \frac{(0.7/12)^2}{11}} = 9$$

وبالكشف في جدول t بدرجات حرية 9 نستخرج قيمة t' - الجدولية كما يلي:

$$t'_{\alpha/2,d.f} = t'_{0.025,9} = 2.26$$

(3) من العينة نحسب الآتي:

$$t' = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} = \frac{25 - 21}{\sqrt{\frac{3.5}{8} + \frac{0.7}{12}}} = 5.71$$

(4) وحيث أن قيمة t' تقع في مناطق الرفض، نرفض الفرض العدم ونقبل البديل.

(4.3.4) اختبار الفرق بين معاملي ارتباط مجتمعين

إذا سحبت عينتان عشوائيتان من مجتمعين معاملات الارتباط فيهما ρ_1 , ρ_2 على التوالي، وكان حجم العينة الأولى n ومعامل الارتباط فيها r_1 وحجم العينة الثانية m ومعامل الارتباط فيها r_2 فان المطلوب باستخدام البيانات المتحصل عليها من العينتين معرفة هل هناك فرق معنوي بين معاملات الارتباط في كل من المجتمعين أم لا، ولإجراء هذا الاختبار يكون لدينا فرض العدم والبديل الآتيين:

$$H_0 : \rho_1 = \rho_2$$
$$H_1 : \rho_1 \neq \rho_2$$

وبقبول فرض العدم يعني أن العينتين مسحوبتان من نفس المجتمع ، وأن القول بأن الفرق بين معاملي الارتباط في المجتمعين غير معنوي ولا يختلفان جوهريا. أما إذا رفض الفرض العدم فإننا نعني بأن معامل الارتباط في المجتمع المسحوبة منه العينة الأولى يختلف عن نظيره المسحوبة منه العينة الثانية. وللحصول على شكل التوزيع العيني الذي يستخدم في هذه الحالة، فقد تم إثبات أن المتغير Z حيث:

$$Z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n-3} + \frac{1}{m-3}}}$$

يخضع للتوزيع المعتدل المعياري إذا كانت حجوم العينات كبيرة، وبذلك يمكن استخدام جداول توزيع Z وكذلك البيانات المستخرجة من العينة لرفض أو قبول الفرض العدم علما بأن:

$$Z_1 = \frac{1}{2}\ln\frac{1+r_1}{1-r_1}$$

$$Z_2 = \frac{1}{2}\ln\frac{1+r_2}{1-r_2}$$

- **مثال (59): من دراسة لعينتين توفرت البيانات الآتية:**

	العينة ألأولى	العينة الثانية
حجم العينة	28	33
معامل الارتباط	0.5	0.7

والمطلوب اختبار ما إذا كانت معاملات الارتباط الخاصة بالمجتمعات المسحوبة منها هذه العينات متساوية أم لا وذلك باحتمال 95%.

الحل:

(1)
$$H_0 : \rho_1 = \rho_2$$
$$H_1 : \rho_1 \neq \rho_2$$

(2) حيث أن الارتباط من طرفين، فانه باستخدام جداول التوزيع المعتدل العياري نجد أن مناطق الرفض والقبول عند $\alpha = 0.05$ هي ± 1.96 .

(3) من العينتين نجد أن قيمة $Z -$ المحسوبة هي:

$$Z_c = \frac{\frac{1}{2}\ln\frac{1+0.5}{1-0.5} - \frac{1}{2}\ln\frac{1+0.7}{1-0.7}}{\sqrt{\frac{1}{28-3}+\frac{1}{33-3}}} = -1.19$$

(4) وحيث أن قيمة $Z -$ المحسوبة تقع في منطقة القبول ، فإننا نقبل H_0 ونجد أن الفرق غير معنوي وان العينتين مسحوبتان من نفس المجتمع.

(4.4) العلاقة بين اختبار الفرضيات وفترات الثقة لتقدير المعالم

إن العلاقة بين تقدير معالم المجتمع والاختبارات الإحصائية للفروض علاقة وثيقة حيث انه يمكن إجراء الاختبارات الإحصائية من خلال تقدير فترة الثقة لمعالم المجتمع الإحصائي مهما كانت نوعية الاختبار الذي يتم اختياره، ذلك لأن حدود الثقة المحددة لتقدير المعلمة μ أو الفرق $\mu_1 - \mu_2$ أو المعلمة p أو الفرق $p_1 - p_2$ ... الخ، هو الإطار المحدد لمنطقتي الرفض والقبول لفرض العدم H_0 بعد تصور ذلك بطريقة معكوسة. فعند اختبار الطرفين للعلاقة بين الإحصائية $\bar{X}$ ومعلمة المجتمع μ (والتي يفترض أنها تأخذ القيمة μ_0) فان الحدين الأدنى والأعلى للقيمة μ_0 هما

$$\bar{X} \mp t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}$$

إذا كان حجم العينة صغيرا . أما إذا كان حجم العينة كبيرا فان الحدين الأدنى والأعلى هما

$$\bar{X} \mp Z_{\alpha/2} \frac{S}{\sqrt{n}}$$

وبالتالي إذا ما وقعت القيمة μ_0 بين الحدين الأعلى والأدنى فإنها تقع في منطقة القبول وهنا يقبل H_0 ويرفض H_1 والعكس صحيح إذا ما وقعت خارج هذه الحدود.

هذا إذا كان الاختبار من طرفين، أما في اختبار الطرف الأيسر فان الحد الأقصى للثقة هو:

$$\bar{X} + t_{\alpha,n-1} \frac{S}{\sqrt{n}}$$

أو

$$\bar{X} + Z_{\alpha} \frac{S}{\sqrt{n}}$$

في العينات الصغيرة والكبيرة على التوالي، فإننا نقبل H_0 عندما تقع μ_0 في منطقة القبول (أقل من حدها الأقصى) والعكس صحيح عندما تقع μ_0 على يمين الحد الأقصى.

وأخيرا ففي حالة اختبار الطرف الأيمن فان الحد الأدنى للثقة (أي قيمة μ_0 كحد أدنى) هو

$$\bar{X} - t_{\alpha,n-1}\frac{S}{\sqrt{n}}$$

أو

$$\bar{X} - Z_{\alpha}\frac{S}{\sqrt{n}}$$

في العينات الصغيرة والكبيرة على التوالي، فإننا نقبل H_0 ونرفض H_1 عندما تأخذ μ_0 قيمة أكبر من حدها الأدنى والعكس صحيح إذا ما وقعت μ_0 في منطقة أبعد من حدها الأدنى.

ويمكن تصور نفس الحدود للفرق بين $\mu_1 - \mu_2$ وكذلك للقيمة p وللفرق $p_1 - p_2$.

وليس من الصعب تفسير ذلك على الإطلاق، فنحن نعلم من اختبارات الفروض إننا نرفض H_0 ونقبل

H_1 عندما تكون t - المحسوبة أكبر من أو تساوي t - الجدولية إذا كان الاختبار من طرف واحد وهو الطرف الأيمن، أي $t_c \geq t_{\alpha,n-1}$، وبالتعويض عن قيمة $t-$ المحسوبة نجد أن:

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq t_{\alpha,n-1}$$

$$\therefore \bar{X} - \mu_0 \geq \frac{S}{\sqrt{n}}(t_{\alpha,n-1})$$

ومنها نجد أن:

$$\mu_0 \le \overline{X} - t_{\alpha,n-1}\frac{S}{\sqrt{n}}$$

وهذا هو شرط رفض H_0 في اختبار الطرف الأيمن. وبنفس الطريقة يمكننا استنتاج حالتي اختبار الطرف الأيسر واختبار الطرفين.

- **مثال (60)**: سحبت عينة عشوائية من مجتمع $N(\mu,\sigma^2)$ تتكون من 10 أكياس من الحليب المجفف حيث أن وزن الكيس الواحد 250 غم بانحراف معياري 3 غم. باستخدام فكرة فترات الثقة المطلوب اختبار الفروض الآتية عند مستوى معنوية 5%:

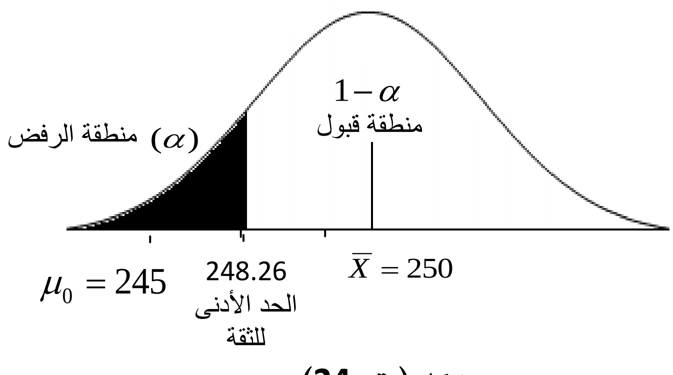
$$H_0 : \mu \le 245$$
$$H_1 : \mu > 245$$

الحل:

نحن هنا بصدد اختبار الطرف الأيمن وبالتالي فان الحد الأدنى للقيمة هي:

$$= \overline{X} - t_{\alpha,n-1}\frac{S}{\sqrt{n}}$$
$$= 250 - t_{0.05,9}\frac{3}{\sqrt{10}}$$
$$= 250 - (1.833)(\frac{3}{\sqrt{10}}) = 248.26$$

وهو الحد الأدنى لقيمة μ_0 . ونظرا لأن $\mu_0 = 245$ (وهي تمثل القيمة الافتراضية للمعلمة μ) أصغر من الحد الأدنى للثقة، أي أنها تقع في منطقة الرفض فإننا نرفض H_0 ونقبل H_1 مما يعني أن هذه الأكياس تعطي وزنا يقل بالفعل عن 245 وهذا ما يوضحه الشكل التالي.



شكل (رقم 24)

- **مثال (61):** في المثال السابق، اختبر الفروض أدناه باستخدام فكرة فترة الثقة وباحتمال 99%.

$$H_0 : \mu = 248$$
$$H_1 : \mu \neq 248$$

الحل: نوجد الحدين الأدنى والأعلى لـ μ_0 عند احتمال 99% كما يلي:

$$LL = \overline{X} - t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}$$
$$= 250 - t_{0.005,9} \frac{3}{\sqrt{10}}$$
$$= 250 - (3.25)(\frac{3}{\sqrt{10}}) = 246.92$$

والحد الأعلى هو:

$$UL = \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$
$$= 250 + t_{0.005,9}\frac{3}{\sqrt{10}}$$
$$= 250 + (3.25)\frac{3}{\sqrt{10}} = 253.08$$

ونظرا لأن قيمة $\mu_0 = 248$ تنحصر بين الحدين الأعلى والأدنى للثقة، فأننا نقبل H_0 ونرفض H_1.

- **مثال (62):** في المثال السابق وباستخدام فكرة فترة الثقة المطلوب اختبار الفروض أدناه عند احتمال 95%.

$$H_0 : \mu \geq 251$$
$$H_1 : \mu < 251$$

الحل: الاختبار هنا هو اختبار الطرف الأيسر وعليه فان الحد الأعلى للثقة هو:

$$UL = \overline{X} + t_{\alpha,n-1}\frac{S}{\sqrt{n}}$$
$$= 250 + t_{0.05,9}\frac{3}{\sqrt{10}}$$
$$= 250 + (1.83)\frac{3}{\sqrt{10}} = 251.74$$

ونظرا لأن قيمة $\mu_0 = 251$ أقل من الحد الأعلى للثقة (251.74) فهي تقع في منطقة القبول وعليه فإننا نقبل H_0 ونرفض H_1 عند $\alpha = 5\%$.

- **مثال (63):** عينة عشوائية تتكون من 8 مفردات وجد فيها أن $\overline{X} = 105$ والانحراف المعياري $S_1 = 4$. ومن عينة أخرى بنفس الحجم وجد أن $\overline{Y} = 100$ بانحراف معياري $S_2 = 4.6$. فإذا كان المجتمعان يتوزعان توزيعا طبيعيا بتباينين مجهولين

ولكنهما متساويان، فان المطلوب باستخدام فكرة فترة الثقة اختبار الفروض أدناه باحتمال 95%:

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$

الحل:

المطلوب هنا إجراء اختبار الطرفين للفرق $\mu_1 - \mu_2$، وبالتالي فان الحد الأدنى والأعلى للثقة هما:

$$LL = (\bar{X} - \bar{Y}) - t_{\alpha/2, n+m-2} \sqrt{S_p^2 (\frac{1}{n} + \frac{1}{m})}$$
$$= (105 - 100) - t_{0.025, 14} \sqrt{S_p^2 (\frac{1}{8} + \frac{1}{8})}$$

وحيث أن: $t_{0.025,14} = 2.14$ وأن: $S_p^2 = \frac{(8-1)(4) + (6-1)(4.6)}{8+8-2} = 4.3$

وعليه يكون الحد الأدنى:

$$LL = 5 - (2.14)\sqrt{4.3(\frac{1}{8} + \frac{1}{8})} = 2.79$$

آما الحد ألأعلى:

$$UL = 5 + (2.14)\sqrt{4.3(\frac{1}{8} + \frac{1}{8})} = 7.21$$

ونظرا لأن قيمة الفرق المقدرة $\mu_1 - \mu_2 = 0$ وهو خارج حدود الثقة، لذلك فإننا نرفض H_0 ونقبل H_1 مما يعني وجود زيادة حقيقية للمتوسط الأول على الثاني.

(5.4) ملخص للاختبارات الإحصائية

المختبر الإحصائي	وصف المجتمع	المعلمة المجهولة
$Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	المجتمع طبيعي. أو $n>30$. σ^2 معلوم.	μ
$t_{n-1}=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	المجتمع طبيعي. $n<30$. σ^2 غير معلوم.	μ
$Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$	المجتمعات طبيعية ومستقلة. أو $n+m>50$. σ_1^2,σ_2^2 معلومة.	$\mu_1-\mu_2$
$t=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}}$	المجتمعات طبيعية ومستقلة. $n+m<50$. σ_1^2,σ_2^2 مجهولة ومتساوية.	$\mu_1-\mu_2$
$t'=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{(\frac{S_1^2}{n}+\frac{S_2^2}{m})}}$	المجتمعات طبيعية ومستقلة. $n+m<50$. σ_1^2,σ_2^2 مجهولة وغير متساوية.	$\mu_1-\mu_2$
$t=\frac{\bar{d}-\mu_d}{S_d/\sqrt{n}}$	المجتمعات طبيعية غير مستقلة.	$\mu_1-\mu_2$

المختبر الإحصائي	وصف المجتمع	المعلمة المجهولة
$Z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	محاولات مستقلة متكررة. n كبيرة ($n>25$).	p
$Z=\frac{X-0.5-np_0}{\sqrt{np_0(1-p_0)}}$	محاولات مستقلة متكررة. n صغيرة ($n<25$).	p
$Z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\tilde{p}\tilde{q}(\frac{1}{n}+\frac{1}{m})}}$	حجم العينات كبير.	$p_1-p_2=0$
$Z=\frac{(\hat{p}_1-\hat{p}_2)-d_0}{\sqrt{(\frac{\hat{p}_1\hat{q}_1}{n}+\frac{\hat{p}_2\hat{q}_2}{m})}}$	حجم العينات كبير.	$p_1-p_2=d_0$
$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$	التوزيع طبيعي.	σ^2
$F_{n-1,m-1}=\frac{S_1^2}{S_2^2}$	التوزيع طبيعي.	σ_1^2/σ_2^2
$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	التوزيع طبيعي.	$\rho=0$
$Z=(Z_1-Z_2)\sqrt{n-3}$	التوزيع طبيعي.	$\rho\neq 0$
$Z=\frac{Z_1-Z_2}{\sqrt{\frac{1}{n-3}+\frac{1}{m-3}}}$	التوزيع طبيعي. $n+m$ كبيرة.	$\rho_1-\rho_2=0$
$t_{n-2}=\frac{\hat{\beta}-\beta_0}{S(\hat{\beta})}$	التوزيع طبيعي.	$\beta=\beta_0$

تمـــارين

(1) في مصنع متخصص بإنتاج البطاريات الكهربائية بعمر 80 ساعة للبطارية الواحدة، سحبت عينة من 20 بطارية وكان متوسط أعمار البطاريات هو 75 ساعة. المطلوب اختبار الفرض القائل أن متوسط أعمار البطاريات هو 80 ساعة ضد الفرض القائل أن المتوسط أقل من 80 ساعة باحتمال 95% علما بأن تباين المجتمع يساوي 41.

(2) سحبت عينة من 15 شخصا من مجتمع طلابي معين فوجد أن 7 أشخاص في هذه العينة مصابون بإدمان التدخين. المطلوب باحتمال 99% اختبار هل أن نسبة الإصابة بإدمان التدخين في ذلك المجتمع هي أكبر من أو يساوي 60%.

(3) في عينة من 49 وحدة وجد أن الانحراف المعياري هو 8. المطلوب باحتمال 95% اختبار الفرض القائل أن هذه العينة مسحوبة من مجتمع انحرافه المعياري يساوي 10.

(4) سحبت عينة مكونة من 10 طلاب في المرحلة الثانوية وقيست أطوالهم فكانت:

120, 100, 130, 140, 145, 150, 122, 145, 150, 155

والمطلوب باحتمال 95% اختبار الفرض القائل أن هذه العينة مسحوبة من مجتمع متوسط الأطوال فيه أقل من 120.

(5) من البيانات الآتية وباحتمال 95% المطلوب:

(1) اختبار هل هناك فرقا معنويا بين متوسطي المجتمعين المسحوبة منها العينتين؟

(2) هل يمكن القول أن: $\mu_1 - \mu_2 < 2$ ؟

	المتوسط	التباين	حجم العينة
العينة الأولى	11.3	7.3	11
العينة الثانية	13.7	10.2	21

(6) توفرت البيانات التالية عن العينتين أدناه:

	العينة الأولى	العينة الثانية
المتوسط	90.3	96.7
حجم العينة	10	20

المطلوب اختبار هل هناك فرقا معنويا بين متوسطي العينتين علما بأن الانحراف المعياري للمجتمع المسحوبة من العينة الأولى 5.3، والانحراف المعياري للمجتمع الآخر المسحوبة منه العينة الثانية 8 وذلك باحتمال 95% ؟

(7) عينتان، الأولى حجمها 10 ونسبة المدخنين فيها 0.30 والثانية حجمها 20 ونسبة المدخنين فيها 0.6، فهل يمكن الحكم أن نسبة المدخنين في المجتمع المسحوبة منه العينة الأولى أكبر من نسبة المدخنين في المجتمع المسحوبة منه العينة الثانية باحتمال 90%؟

(8) من البيانات الآتية: اختبر التجانس بين المجموعتين ثم قدر تباين المجتمعين إن أمكن وذلك باحتمال 95%.

	العينة الأولى	العينة الثانية
التباين	55.7	36.6
حجم العينة	45	35

(9) أخذت عينتان، حجم الأولى 100 والثانية 200 من توزيعين أوساطهم μ_1, μ_2 وتبايناتهم σ_1^2, σ_2^2 على التوالي. درست هاتان العينتان ووجد أن متوسط العينة

الأولى وتباينها هما: 81, 50 على التوالي، وأن متوسط العينة الثانية وتباينها هما على التوالي: 60 ,64. والمطلوب اختبار الفرضيات الآتية عند مستوى دلالة 0.01:

(أ) $H_0 : \mu_1 = 55$ مقابل الفرض $H_a : \mu_1 < 55$

(ب) $H_0 : \mu_1 = \mu_2$ مقابل الفرض $H_a : \mu_1 < \mu_2$

(ج) $H_0 : \sigma_2^2 = 71$ مقابل الفرض $H_a : \sigma_2^2 < 71$

(د) $H_0 : \sigma_1^2 = \sigma_2^2$ مقابل الفرض $H_a : \sigma_1^2 > \sigma_2^2$

(10) في السؤال السابق، المطلوب اختبار الفرضيات أ ، ج باستخدام العلاقة بين اختبار الفرضيات وفترات الثقة لتقدير المعالم.

(11) سحبت عينة عشوائية من 5 مفردات من مجتمع وقيست فيها ظاهرتان لمتغير مستقل *X* وآخر تابع *Y* فكانت النتائج كالآتي:

X	5	4	3	2	1
Y	2	4	6	3	4

والمطلوب: (أ) إيجاد معامل الارتباط ومعامل الانحدار.

(ب) إيجاد فترة ثقة 95% لكل من معامل الانحدار ومعامل الارتباط.

(ج) اختبر الفرضية $H_0 : \rho = 0$ على مستوى دلالة 5%.

(د) اختبر الفرضية $H_0 : \rho = 0.55$ على مستوى دلالة 5%.

(هـ) اختبر الفرض $H_0 : \beta = 0$ على مستوى دلالة 1%.

(12) توفرت البيانات التالية عن عينتين مسحوبتين من مجتمعين طبيعيين:

العينة	المتوسط	التباين
الأولى	12	?
الثانية	10	9

فإذا علم أن التباين المجمع S_p^2 يساوي 15، فإن المطلوب باحتمال 99%:

(1) اختبار أن الفرق بين متوسط المجتمع الأول والثاني لا يتعدى 2.

(2) فترة ثقة لتباين العينة الأولى.

(13) إذا توفرت لديك المجاميع الآتية، فإن المطلوب إيجاد فترة الثقة المطلوبة:

$$(a)\ \sum (X-\overline{X})^2 = 32, \quad \sum X = 450, \quad n=9, \quad 1-\alpha = 95\%$$

$$(b)\ \sigma^2 = 64, \quad \overline{X} = 52, \quad n=16, \quad 1-\alpha = 99\%$$

(14) توفرت البيانات الآتية عن نوعين من إطارات السيارات:

بيانات تتعلق بالنوع الأول من الإطارات	بيانات تتعلق بالنوع الثاني من الإطارات
$n = 21$ $\overline{X} = 26000$ $S_1 = 1100$	$m = 25$ $\overline{Y} = 13200$ $S_2 = 1500$

والمطلوب اختبار الفرض القائل أن متوسط عمر إطارات النوع الأول يزيد على الأكثر بـ 10000 كم عن متوسط عمر إطارات النوع الثاني.

(15) لاختبار تأثير علاج معين على ضغط الدم، قيس الضغط لـ 12 فردا قبل وبعد أخذ العلاج، فكانت الزيادة في ضغط دمهم كالتالي:

4, 1, 7, 0, 2, -1, -3, 0, 4, -1, 3

والمطلوب: (1) إيجاد فترة ثقة حول متوسط تغير ضغط الدم عند مستوى احتمال 5%.

(2) اختبر الفرض القائل أن $\mu_d = 0$.

(16) أذكر كلا من: الفرض البديل، دالة الاختبار، القيمة الجدولية، ونوع الاختبار لكل فرض من الفروض التالية:

$$(a)\ H_0 : \sigma_1^2 = \sigma_2^2$$

$$(b)\ H_0 : \sigma^2 \geq 13$$

$$(c)\ H_0 : p_1 \leq p_2$$

$(d)\ H_0 :$ البيانات تتبع التوزيع ألأسي

الفصل الخامس

مربع كاي (χ^2) واختبارات حسن المطابقة

(1.5) مقدمة.

(2.5) اختبار χ^2 .

(3.5) اختبارات الاستقلالية.

(4.5) حسن المطابقة.

(5.5) اختبار كولومجروف.

كاي تربيع (χ^2) واختبارات جودة التوفيق

Chi Square and Goodness of Fit Tests

(1.5) مقدمة (*Introduction*)

كانت الاستخدامات المختلفة لكل من توزيع Z وتوزيع t تتركز في تقدير بعض معالم المجتمعات الإحصائية مثل المعلمة μ أو p أو الفرق $\mu_1 - \mu_2$ أو $p_1 - p_2$ ، إلى جانب إجراء بعض الاختبارات الإحصائية لعلاقة الإحصائية $\overline{X}$ بالمعلمة μ أو اختبارات المعنوية المتعلقة بالفرق $\mu_1 - \mu_2$ سواء في ضل استخدام العينات الكبيرة أو الصغيرة.

غير أن التوزيعات المذكورة قد لا تساعد إلى درجة كبيرة في اختبار العلاقة بين ما هو متاح للباحث من معلومات مشاهدة عن ظاهرة معينة وما يمكن حسابه من معلومات نظرية من خلال قانون رياضي معين يحكم ويحرك هذه الظاهرة. وهنا يلجأ الباحث إلى استخدام توزيع كاي تربيع (χ^2) لإجراء مثل هذا الاختبار لمعرفة معنوية العلاقة بين البيانات المشاهدة والبيانات المتوقعة من خلال قانون رياضي معروف. وتتسع المجالات المختلفة لإجراء اختبار(χ^2) لتشمل الحالات التي نجد فيها أن المتغير الإحصائي محل الدراسة يأخذ بيانات وصفية لا كمية حيث يتعذر على الاختبارات السابقة المساعدة في مثل هذا الاختبار. والمتغير الوصفي هو المتغير الغير قابل للقياس الكمي مثل صفة التدخين أو الأمية أو المرض.

كما يمكن استخدام اختبار χ^2 في دراسة معنوية العلاقة بين بعض البيانات الكمية وأخرى وصفية، كدراسة العلاقة المعنوية نتيجة إعطاء مصل معين ضد الإصابة بأحد

الأمراض والتي يمكن صياغتها في أحد جداول الاقتران أو التوافق حيث يتم تقسيم مفردات العينة إلى (مجموعة أخذت المصل وأخرى لم تأخذه) وتقسيم النتيجة أيضا إلى (مجموعة لم تصب بالمرض وأخرى أصيبت به).

ومن أهم مجالات اختبار χ^2 التطبيقية هو علم الوراثة والطبيعة والزراعة والكيمياء، حيث يتم اختبار صحة بعض النظريات. ففي علم الوراثة مثلا قد يريد الباحث معرفة ما إذا كانت بيانات معينة تتفق مع فروض نظرية مشهورة كنظرية مندل مثلا. إلى جانب هذه المجالات المختلفة لاختبار χ^2 فإنه يستخدم في اختبار جودة توفيق بعض المنحنيات إلى المنحنى المعتدل أو إلى منحنى بواسون أو إلى منحنى ذي الحدين، بمعنى آخر اختبار مدى تبعية بعض المشاهدات المتاحة لتوزيع احتمالي له دالة احتمالية معروفة. فقد نقول أن أعداد حوادث السير التي تحدث عند تقاطع طرق معين تخضع لتوزيع بواسون، وقد نقول أن عدد الوحدات المعيبة المنتجة في مصنع معين تخضع لتوزيع ذات حدين، ولكن السؤال في حالات كهذه هو كيف يمكن التأكد من أن المشاهدات المتوفرة تخضع لهذا النموذج الإحصائي أو ذاك؟ أي أننا نريد اختبار مدى جودة التوفيق أو مدى حسن مطابقة النموذج الإحصائي للبيانات المتوفرة. ولا نريد أن ننسى استخدامنا لـ χ^2 في اختبار معنوية التباين σ^2 وإيجاد فترات الثقة له.

وخلاصة القول فإن اختبارات χ^2 تتركز في ثلاث اتجاهات:

(1) اختبار صحة بعض النظريات من خلال البيانات المشاهدة عن ظاهرة تخضع لقانون محدد.

(2) اختبارات الاستقلالية في جداول الاقتران والتوافق.

(3) اختبارات جودة التوفيق.

(2.5) اختبار χ^2 (*Chi Square Test*)

لنفرض أن لدينا ظاهرة معينة، ولكي نتأكد أن هذه الظاهرة تخضع لنظرية معينة نسحب عينة عشوائية حجمها n، ومن هذه العينة نجمع قيما أو معلومات نطلق عليها

اسم القيم المشاهدة O_i (*Observed Values*)، ثم وباستخدام النظرية يمكن الحصول على قيم أخرى نطلق عليها اسم القيم النظرية أو المتوقعة e_i (*Expected Values*)، وأخيرا وباستخدامنا لاختبارات الفروض يمكن التأكد من وجود أم عدم وجود فروق معنوية بين القيم المشاهدة للظاهرة والقيم المتوقعة حيث تقارن التكرارات أو القيم النظرية المتوقعة والتي نحصل عليها نتيجة توفيق بعض المنحنيات كالمنحنى الطبيعي أو منحنى بواسون بالتكرارات الفعلية المشاهدة والتي نحصل عليها من واقع المشاهدة أو القياس المباشر، نجد أن هناك فروقا بين التكرارات المتناظرة والتي لابد من حدوثها، لأن التوزيع التكراري المشاهد يمثل عينة محدودة في حين أن التوزيع التكراري النظري يصف المجموع ولا يمكن أن تكون العينة مطابقة تماما للمجموع الذي اختيرت منه. غير أن حجم هذه الفروق قد يكون من الكبر، بحيث يثبت عدم صلاحية المنحنى (الذي وفقناه للتوزيع) لوصف المجموع الأصلي.

وأول طريقة تتبادر إلى الذهن لاختبار حسن مطابقة التوزيع النظري الذي حصلنا عليه بتوفيق أحد المنحنيات للتوزيع المشاهد هي أن نعين الفروق (بعد إهمال إشاراتها) بين التكرارات المشاهدة والتكرارات النظرية المقابلة لها ثم نحسب مجموعها، وكلما كان هذا المجموع كبيرا كلما دل ذلك على عدم صلاحية المنحنى الذي وفقناه لوصف تغير الظاهرة تحت البحث.

غير أن مجرد جمع الفروق مع إهمال إشاراتها لا يكفي لإعطائنا مقياسا دقيقا لاختبار جودة التوفيق للمنحنى الذي وفق للتوزيع المشاهد، إذ أن من الواضح أن هذا المقياس يتوقف إلى حد كبير على عدد التكرارات. فإذا كان هذا العدد كبيرا كان مجموع الفروق المطلقة بين التكرارات المشاهدة والمتوقعة كبيرا حتى وأن كانت المطابقة جيدة وملائمة. ولكي نتغلب على هذه الصعوبة نحسب مجموع الفروق كنسبة مئوية من مجموع التكرارات لنحصل على مقياس يسمى " النسبة المئوية لسوء المطابقة" **والذي يعبر عنه رمزيا كما يلي:**

$$\frac{\sum|O_i - e_i|}{\sum O_i} \times 100$$

حيث O_i تمثل التكرارات المشاهدة، e_i تمثل التكرارات المتوقعة أو النظرية. ويعاب على هذا المقياس انه لا يعطينا حدا فاصلا بين النسبة التي نعتبرها كبيرة (والتي تدل على عدم جودة التوفيق) بل أن الأمر متروك لتقدير الباحث الشخصي.

ولكن الطريقة الأكثر دقة هي التي وضعها بيرسون المعروفة باختبار χ^2 لاستخدامها في الاتجاهات المذكورة سابقا. ويمتاز هذا الاختبار على المقياس السابق في انه يحدد لنا مستوى موضوعيا نحكم بمقتضاه على ما إذا كانت الفروق التي نجدها بين التوزيع المشاهد والتوزيع النظري (أي بين التكرارات المشاهدة والمتوقعة) فروقا جوهرية تدل على عدم صلاحية النموذج الإحصائي أو المنحنى الموفق لوصف المجموع الذي انتخبت منه القيم المشاهدة، أو فروقا ظاهرية يمكن أن تعزى إلى عوامل المصادفة فلا تدل على وجود اختلاف أساسي بين التوزيعين النظري والمشاهد. ويتوقف هذا الاختبار على حساب قيمة χ^2 التي تعرف كالآتي:

$$\chi_c^2 = \sum_i \frac{(O_i - e_i)^2}{e_i}$$

و فيما يلي خطوات العمل اللازمة لحساب قيمة χ^2 :

(1) نعين انحرافات التكرارات المشاهدة عن التكرارات المتوقعة بطرح الثانية من الأولى، أي نوجد قيمة $O_i - e_i$.

(2) حيث أن جمع هذه الانحرافات جمعا جبريا لا يعطينا فكرة صحيحة عن حجم الفروق بين التكرارات المشاهدة والتكرارات المتوقعة، كما سبق وأن ذكرنا، فإننا نتخلص من إشارة الانحرافات بتربيعها بدلا من مجرد إهمالها. وعلى ذلك تكون الخطوة الثانية هي تعيين قيم $(O_i - e_i)^2$.

(3) نقسم مربع كل انحراف على التكرار المتوقع المقابل له لبيان أهميته النسبية، وذلك لأن الانحراف قد يكون كبيرا أو صغيرا لمجرد أن التكرار نفسه كبير أو صغير. وعلى ذلك فإن الخطوة الثالثة هي إيجاد قيم $(O_i - e_i)^2 / e_i$.

(4) نجمع قيم $(O_i - e_i)^2 / e_i$ ، و المجموع الذي نحصل عليه هو قيمة χ^2 .

ومن الواضح أنه كلما كانت قيمة χ^2 كبيرة كلما دل ذلك على أن الفروق بين التكرارات النظرية والمشاهدة كبيرة. ويتم اختبار χ^2 بإجراء الخطوات السابقة لاختبارات الفروض وهي:

(أ) تحديد الفرض العدم والفرض البديل

القاعدة العامة عند وضع الفروض الإحصائية المتعلقة باختبار χ^2 أنه لا يوجد فرق معنوي حقيقي بين البيانات المستمدة من المشاهدات والبيانات النظرية التقديرية المستمدة من واقع نظرية معينة أو قاعدة رياضية معروفة أو قانون وراثي ثابت وذلك في مقابل الفرض البديل بوجود اختلاف بينهما، أي أن:

H_0 : النظرية صحيحة أو القاعدة صحيحة أو الصفات مستقلة أو التوفيق جيد.

H_1 : النظرية غير صحيحة أو القاعدة غير صحيحة أو الصفات غير مستقلة أو التوفيق غير جيد.

(ب) تم إثبات أن المتغير χ^2 حيث:

$$\chi^2 = \sum_i \frac{(O_i - e_i)^2}{e_i} = \sum_1 \frac{O_i^2}{e_i} - N$$

يتبع توزيع χ^2 بدرجات حرية n-1.

(ج) باستخدام العينة يمكن حساب قيمة $\chi^2 -$ المحسوبة فإذا كان التكرار المشاهد يساوي التكرار المتوقع فإن قيمة χ^2 تساوي صفرا وتزداد قيمة χ^2 بزيادة الفرق بين القيم المشاهدة والمتوقعة. وبمقارنة قيمة $\chi^2 -$ المحسوبة بقيمة $\chi^2 -$ الجدولية باحتمال معين يمكن استنتاج الآتي:

(1) إذا كانت $\chi_c^2 < \chi_{n-1}^2$ فإن الفرق بين القيم المشاهدة O_i والقيم المتوقعة e_i فرق غير معنوي وعليه نقبل الفرض العدم H_0 .

(2) إذا كانت $\chi_c^2 \geq \chi_{n-1}^2$ يكون الفرق بين التكرار المشاهد والمتوقع فرقا معنويا وهنا نقبل الفرض البديل.

مع ملاحظة أن الاختبار هو اختبار من طرف واحد دائما وهو اختبار الطرف الأيمن.

- **مثال (64)**: ألقينا قطعة نقود 200 مرة فحصلنا على 114 وجه وعلى 86 ظهر، فهل ترى أن هذه العملة متزنة باحتمال 95%؟

الحل:

إذا كانت هذه العملة متوازنة فإننا نتوقع الحصول على 50% صورة من عدد المرات و 50% كتابة، وعليه يمكن إجراء الاختبار كما يلي:

(1) $H_0:$ العملة متوازنة

$H_1:$ العملة غير متوازنة

(2) حيث أن التوزيع الذي سوف يستخدم هو توزيع χ^2 فإنه باستخدام جداول χ^2 تكون القيمة الحرجة التي على أساسها يقبل الفرض العدم أو الفرض البديل هي:

$$\chi_{\alpha,d.f}^2 = \chi_{0.05,1}^2 = 3.84$$

حيث 1 تمثل درجات الحرية والتي تساوي:

درجات الحرية = عدد المتغيرات − 1

= 2 - 1 = 1

(3) يمكن حساب χ^2 – المحسوبة كالتالي: حيث أننا نتوقع الحصول على 50% من عدد المرات على شكل صورة (أي نصف عدد المرات) وكذلك الحال بالنسبة للظهر، فإن التكرار المتوقع في هذه الحالة هو 100 مرة وجه و100 مرة ظهر. وبالتالي فإن قيمة χ^2 سوف تكون:

$$\chi_c^2 = \sum_i \frac{(O_i - e_i)^2}{e_i} = \frac{(114-100)^2}{100} + \frac{(86-100)^2}{100} = 3.92$$

(4) وحيث أن قيمة χ^2 - المحسوبة أكبر من χ^2 - الجدولية لذلك يرفض الفرض العدم ويقبل الفرض البديل القائل بأن العملة متحيزة وغير متوازنة.

- **مثال (65):** في تجربة لتربية نوع من أنواع الزهور، فحص مهندس زراعي 400 نوعا من أنواع البذور المختلفة فكانت النتائج الآتية:

أنواع البذور	التكرار المشاهد
بذور مستطيلة صفراء	200
بذور غير مستطيلة صفراء	60
بذور مستطيلة خضراء	60
بذور غير مستطيلة خضراء	80

والمطلوب اختبار ما إذا كانت هذه البيانات تتطابق مع فروض النظرية التي تقول أن نسب توزيع البذور هي 2 :1 :1 :1 وذلك باحتمال 95%.

الحل:

(1) $H_0:$ النظرية صحيحة

$H_1:$ النظرية غير صحيحة

(2) القيمة الحرجة التي على أساسها سيقبل فرض العدم باستخدام جداول χ^2 هي:

$$\chi^2_{\alpha,d.f} = \chi^2_{0.05,4-1} = 7.81$$

(3) طبقا للنظرية فإن التكرارات المتوقعة تتوزع على الشكل التالي:

حيث أن النسب هي 2 :1 :1 :1 فإن مجموع الأجزاء هي 5 ويكون التكرار المتوقع كما يلي:

التكرار المتوقع (e_i)	نوع البذور
$e_1 = N(\frac{2}{5}) = 400(\frac{2}{5}) = 160$	بذور مستطيلة صفراء
$e_2 = 400(\frac{1}{5}) = 80$	بذور غير مستطيلة صفراء
$e_3 = 400(\frac{1}{5}) = 80$	بذور مستطيلة خضراء
$e_4 = 400(\frac{1}{5}) = 80$	بذور غير مستطيلة خضراء

بعد ذلك يمكن أن نصور جدول χ^2 كما يلي:

جدول (رقم 23)

نوع البذور	O_i	e_i	$(O_i - e_i)$	$(O_i - e_i)^2$	$\frac{(O_i - e_i)^2}{e_i}$
مستطيلة صفراء	200	160	40	1600	10
غير مستطيلة صفراء	60	80	-20	400	5
مستطيلة خضراء	60	80	-20	400	5
غير مستطيلة خضراء	80	80	0	0	0
المجموع	**400**	**400**			**20**

(4) نظرا لأن χ^2 – المحسوبة اكبر من قيمة χ^2 – الجدولية، نقبل الفرض البديل الذي يقول أن هذه البيانات لا تخضع لهذه النظرية، أو أن النظرية غير صحيحة.

(3.5) اختبارات الاستقلالية (*Tests of Independence*)

في مسائل علمية كثيرة، نقوم بتقسيم أو تصنيف مجموعة من البيانات وفق أسلوبين. فقد نصنف أفراد مجتمع ما وفق خاصية التعليم من جهة ووفق خاصية التدخين من جهة أخرى. وفي مثل هذه الحالات عادة ينشأ السؤال التالي: هل هناك علاقة بين أسلوبي التقسيم؟ بمعنى هل هناك علاقة بين التدخين والتعليم؟ وللإجابة على هذا السؤال نقوم بإجراء اختبار الاستقلال وذلك باستخدام اختبار χ^2 على أن نفرق بين طريقة إجراء الاختبار في حالة جداول الاقتران وجداول التوافق.

أولا: جداول الاقتران (*Association Table*)

وهو عبارة عن جدول تكراري مزدوج (2×2) ، أي أن كل خاصية أو كل صفة في العينة لا تنقسم إلا إلى قسمين فقط مع وجود صفتين لا غير في دراسة العلاقة المعنوية. فإذا أردنا أن ندرس العلاقة بين التدخين والتعليم فإن جدول الاقتران سوف يكون على الشكل التالي.

جدول اقتران (2×2)

تدخين / تعليم	نعم	لا	المجموع
نعم	*a*	*B*	n_1
لا	*c*	*D*	n_2
المجموع	n_3	n_4	*N*

ولإجراء هذا الاختبار نستخدم اختبار χ^2 كما يلي:

(1) الصفات مستقلة $H_0:$

الصفات غير مستقلة $H_1:$

(2) نستخرج χ^2 من الجدول باحتمال α وبدرجات حرية هي كالآتي:

درجات الحرية = (عدد الصفوف – 1) (عدد الأعمدة – 1) = 1

(3) تحسب χ^2 مباشرة من البيانات كما يلي:

$$\chi_c^2 = \frac{N(ad-bc)^2}{(n_1)(n_2)(n_3)(n_4)}$$

حيث أن كلا من *a, b, c, d* تمثل التكرارات المشاهدة في جدول الاقتران السابق.

(4) بمقارنة χ^2 – المحسوبة بالقيمة الجدولية يمكن رفض أو قبول الفرض العدم.

- **مثال (66)**: من عينة عشوائية من مائة فرد وجد أن من بين 70 فردا متعلما هناك 15 فردا يقبل على التدخين. ومن بين الأفراد غير المتعلمين وجد أن 4 أفراد ليس لهم رغبة في الإقبال على التدخين. فهل يمكن إرجاع صفة عدم الإقبال على التدخين أصلا إلى صفة التعليم باحتمال 95%؟

الحل: من بيانات المثال يمكن عمل جدول الاقتران التالي.

تعليم / تدخين	متعلم	غير متعلم	المجموع
يدخن	15	26	**41**
لا يدخن	55	4	**59**
المجموع	**70**	**30**	**100**

ثم نقوم بإجراء الاختبار كالآتي:

(1) صفة التعليم مستقلة عن صفة التدخين $H_0:$

صفة التعليم غير مستقلة عن صفة التدخين $H_1:$

(2) باستخدام جداول χ^2، نجد أن قيمة $\chi^2 -$ الجدولية هي:

$$\chi^2_{\alpha,d.f} = \chi^2_{0.05,(2-1)(2-1)} = \chi^2_{0.05,1} = 3.84$$

(3) تحسب χ^2 من البيانات كما يلي:

$$\chi^2_c = \frac{N(ad-bc)^2}{n_1.n_2.n_3.n_4} = \frac{100((15)(4)-(26)(55))^2}{(41)(59)(30)(70)} = 36.95$$

(4) لأن $\chi^2 -$ المحسوبة أكبر من مثيلتها الجدولية، نرفض H_0 ونقبل H_1، الأمر الذي يعني عدم استقلال الصفات، أي أن صفة التعليم لها علاقة جوهرية ومعنوية مع التدخين.

ثانيا: جدول التوافق (*Contingency Table*)

إذا رجعنا إلى جدول الاقتران السابق وقمنا بتقسيم التعليم إلى أكثر من قسمين فإننا نحصل على جدول التوافق المبين أدناه. وليس من الضروري في مثل هذا النوع من الجداول (أي جداول التوافق) تساوي عدد الصفوف وعدد الأعمدة، فقد نجد جدولا من الرتبة (3×2) أو الرتبة (4×3) أو الرتبة (3×3) ... وهكذا.

جدول توافق (2×3)

تدخين / تعليم	نعم	لا	المجموع
أمي	a	b	n_1
يقرأ	c	d	n_2
متعلم	e	f	n_3
المجموع	n_4	n_5	N

فإذا أردنا أن نفرض هل أن الظاهرتين مستقلتان أم لا، فإننا نجري اختبار χ^2 كالتالي:

(1) $H_0:$ استقلال الصفات

$H_1:$ عدم استقلال الصفات

(2) إيجاد χ^2 − الجدولية باحتمال α وبدرجات حرية تساوي:

درجات الحرية = (عدد الصفوف − 1)(عدد الأعمدة − 1)

(3) لإيجاد χ^2 − المحسوبة نوجد أولا التكرار المتوقع لكل خلية ثم نحسب χ^2 بصيغتها المعروفة. ويحسب التكرار المتوقع لكل خلية على أساس (مجموع تكرارات الصف × مجموع تكرارات العمود) عند مستوى كل خلية ثم ينسب الناتج إلى المجموع الكلي للتكرارات، أي أن:

التكرار المتوقع للخلية الأولى في الصف الأول e_1 يساوي:

$$e_1 = \frac{(n_1)(n_4)}{N}$$

وللخلية الثانية في الصف الأول e_2 يساوي:

$$e_2 = \frac{(n_1)(n_5)}{N}$$

وهكذا بالنسبة لبقية الخلايا.

(4) تقارن χ^2 − المحسوبة بقيمة χ^2 من الجدول ثم نقبل أو نرفض فرض العدم على أساس المقارنة.

- **مثال (67)**: لدراسة تأثير ثلاثة أغذية معينة على مقاومة مرض معين، أخذت ثلاثة مجموعات من الفئران وأطعمت كل مجموعة نوعا معينا من هذه الأغذية، ثم حقنت بمكروب المرض ورصدت أعداد الوفيات والأحياء في كل مجموعة فكانت النتائج الآتية:

	1	2	3	المجموع
أحياء	20	20	10	$n_1 = 50$
وفيات	30	10	30	$n_2 = 70$
المجموع	$n_3 = 50$	$n_4 = 30$	$n_5 = 40$	120

والمطلوب اختبار تأثير الأغذية على مقاومة المرض باحتمال 99%.

الحل:

(1) H_0 : لا توجد علاقة

H_1 : توجد علاقة

(2)
$$d.f = (r-1)(c-1)$$
$$= (2-1)(3-1) = 2$$

ومنها نجد أن:

$$\chi^2_{0.05,2} = 9.21$$

(3) لحساب χ^2 – المحسوبة نقوم بحساب التكرار المتوقع كالتالي:

التكرار المتوقع للخلية الأولى في الصف الأول e_1 هو:

$$e_1 = \frac{(n_1)(n_3)}{N} = \frac{(50)(50)}{120} = 20.83$$

التكرار المتوقع للخلية الثانية في الصف الأول e_2 هو:

$$e_2 = \frac{(n_1)(n_4)}{N} = 12.5$$

التكرار المتوقع للخلية الأولى في الصف الثاني e_4 هو:

$$e_4 = \frac{(n_2)(n_3)}{N} = \frac{(70)(50)}{120} = 29.17$$

التكرار المتوقع للخلية الثانية في الصف الثاني e_5 هو:

$$e_5 = \frac{(n_2)(n_4)}{N} = \frac{(70)(30)}{120} = 17.5$$

وسوف نكتفي بحساب هذه التكرارات ويمكن استنتاج بقية التكرارات المتوقعة كما يلي:

بالنسبة للصف ألأول وجدنا أن التكرار المتوقع للخلية الأولى هو 20.83 وللخلية الثانية هو 12.5، وحيث أن مجموع التكرارات المشاهدة في أي صف أو أي عمود يجب أن يكون مساويا لمجموع التكرارات المتوقعة لذلك فان التكرار المتوقع للخلية الثالثة في الصف الأول e_3 هو:

$$e_3 = 50 - (20.83+12.5)$$
$$= 16.67$$

وهكذا بقية التكرارات. ومن خلال قيم التكرارات المتوقعة يمكن تصوير جدول التكرارات المتوقعة على النحو التالي.

جدول (رقم 24)

	1	2	3	**المجموع**
أحياء	20.83	12.5	16.67	50
وفيات	29.17	17.5	23.33	70
المجموع	**50**	**30**	**40**	**120**

لاحظ عدم حدوث أي تغير في مجاميع ألأعمدة والصفوف. ومن الجدول يمكن إيجاد χ^2 - المحسوبة كالآتي:

$$\chi_c^2 = \sum_i \frac{O_i^2}{e_i} - N$$
$$= \left[\frac{(20)^2}{20.83} + \frac{(20)^2}{12.5} + \cdots + \frac{(30)^2}{23.33}\right] - 120$$
$$= (19.2 + 32 + 6 + 30.85 + 5.71 + 38.58) - 120$$
$$= 12.34$$

(4) نظرا لأن χ^2 – المحسوبة أكبر من χ^2 – الجدولية فإننا نرفض H_0 القائل بأنه لا توجد علاقة بين المتغيرين ونقبل البديل H_1. وبنفس الأسلوب يمكن إجراء هذا الاختبار لأي عدد من الصفوف أو ألأعمدة.

(4.5) حسن المطابقة (*Goodness of Fit*)

تمكنا في دراستنا السابقة من الحصول على بعض التوزيعات التجريبية، وعلمنا كيف يشتق توزيع تجريبي من تجارب عملية كتوزيع ذي الحدين وتوزيع بواسون. وقد تستوفي هذه التجارب خصائص التوزيع وقد لا تستوفي، بمعنى إننا نريد التأكد أن هذا التوزيع الذي تم حسابه من بيانات العينة يتبع توزيعا نظريا معينا أم لا، وبهذا سوف **نقسم الدراسة إلى شقين:**

<u>الأول</u>: هو كيف نوفق توزيعا باستخدام بيانات عينة.

<u>الثاني</u>: هو كيف نختبر جودة هذا التوفيق أو حسن المطابقة.

والشق الثاني يتم باستخدام توزيع χ^2 (راجع البند 2.5)، فباستخدام التوزيع الموفق يمكن الحصول على تكرارات مشاهدة ثم باستخدام التوزيع النظري يمكن الحصول على التكرار المتوقع. وما دام هناك تكرار مشاهد وتكرار متوقع يمكن استخدام الأساليب السابقة وإجراء اختبار χ^2 لمعرفة هل هناك فرق بين التوزيعين أم لا ؟ **وقبل إجراء هذا الاختبار هناك شرطان يجب تحقيقهما في هذه ألأحوال:**

(1) يجب أن يكون مجموع التكرارات كبيرا (أكبر من 30).

(2) يجب أن يكون التكرار المتوقع لكل خلية أكبر من أو يساوي 5 . فإذا كانت $e_i < 5$ لبعض قيم i فإننا ندمج الفئة التي تكرارها أقل من 5 مع الفئة السابقة لها أو الفئة اللاحقة لها. والسبب في ذلك انه إذا كان عدد التكرارات في الفئة صغيرا جدا ، مثل 1 أو 2 أو 3 أو 4، فان دلالة الفرق بين التكرار المشاهد والمتوقع تكون ضعيفة ولا يمكن الاعتماد عليها في الوصول إلى أية نتيجة. وهذا ما يجب

أخذه عند حساب درجات الحرية حيث نأخذ عدد الفئات بعد عملية الضم لتصبح درجات الحرية بالشكل التالي:

درجات الحرية= عدد الفئات أو الصفوف (بعد الضم إن وجد) -عدد المعالم المقدرة -1

وباستخدام الرموز فان: $d.f = r - 1 - k$

حيث:

r = عدد الصفوف في الجدول الناتج بعد عملية الضم.

k = عدد المعالم التي تم تقديرها أو التي تم استخدامها في عملية التوفيق من البيانات المشاهدة.

وفي حالة عدم تحقيق هذين الشرطين فسوف تكون أحكامنا متعرضة لنسبة خطأ كبيرة إذا **لم نتبع أحد حلين**:

<u>الأول</u>: تطويع البيانات بحيث تحقق الشروط لنتمكن من تطبيق χ^2 .

<u>الثاني</u>: استخدام اختبار آخر لا يتطلب شروطا كهذه مثل اختبار كولمجروف، والذي سوف نتعرض له لاحقا.

وفي هذا البند سوف نتعرض لمجموعة من الأمثلة، حيث يناقش كل مثال مشكلة توفيق أحد التوزيعات ثم اختبار جودة التوفيق.

(أ) توفيق منحنى بواسون

إذا كان لدينا توزيع تكراري يمثل حدثا نادر الوقوع غير انه يتكرر في شيء من الانتظام من آن لآخر مثل حادث خروج قطار عن القضبان، أو أي حدث آخر احتمال وقوعه ضئيل جدا، فان أنسب منحنى لوصف التوزيع التكراري هو منحنى بواسون. ويمكن توفيق منحنى بواسون بالتعويض مباشرة في معادلته عن قيمة الوسط الحسابي التي نستخرجها من التوزيع المشاهد. ولا ينسى القارئ أن معادلة هذا المنحنى هي:

$$p(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0,1,2,...$$

حيث λ تمثل الوسط الحسابي للتوزيع وأن x تمثل عدد مرات النجاح (أي قيمة المتغير)، وبالتالي فانه يمكن تعيين التكرارات المتوقعة بمجرد تعيين λ . وفيما يلي نورد مثالا نوضح فيه طريقة توفيق منحنى بواسون.

- **مثال (68):** قام أحد العاملين بميناء المعقل والمسئول عن أسباب انتظار السفن إثناء عملية التفريغ بمشاهدة 200 سفينة إثناء حركة الميناء وسجل البيانات الآتية:

ساعات الانتظار	0	1	2	3	4	5	6
عدد السفن	16	28	107	42	4	2	1

والمطلوب: (1) توفيق توزيع بواسون.

(2) اختبار جودة التوفيق.

الحل:

لتوفيق منحنى بواسون للتوزيع التكراري بالجدول السابق، نبدأ أولا بإيجاد قيمة الوسط الحسابي للتوزيع λ . وبالتعويض عن قيمة λ في المعادلة يمكننا أن نحسب التكرارات المتوقعة مع العلم بأن دالة توزيع بواسون تعطينا التكرارات النسبية التي مجموعها واحد. ولكي نحصل على التكرارات المطلقة نضرب هذه التكرارات النسبية في مجموع التكرارات N. وفيما يلي طريقة العمل.

أولا : توفيق توزيع بواسون:

ذكرنا أن دالة توزيع بواسون هي:

$$p(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0,1,2,...$$

ومرة أخرى إذا أمكن تقدير قيمة λ فإننا سوف نحصل على المنحنى الموفق باستخدام البيانات المتوفرة لدينا. ولتقدير هذه القيمة فان من المعروف أن متوسط التوزيع هو λ وعلى ذلك فان التقدير الغير متحيز للمعلمة λ يساوي الوسط الحسابي لعدد ساعات انتظار السفينة الواحدة، أي أن:

$$\hat{\lambda} = \frac{\sum Xf}{\sum f} = \frac{0(16)+1(28)+\ldots+6(1)}{200} = 2$$

وبالتالي تكون الدالة المتوقعة هي:

$$p(X=x)=\frac{e^{-2}2^{x}}{x!}, \qquad x=0,1,2,...,6$$

ومن هذه الدالة نحصل على التكرار المتوقع كالتالي:

$$e_i = N\,p(X=x)$$

وبتطبيق دالة احتمال توزيع بواسون نوجد التكرارات المتوقعة كالآتي:

$$e_0 = N\,p(X=0)=200(\frac{e^{-2}2^{0}}{0!})=27.07$$

$$e_1 = N\,p(X=1)=200(\frac{e^{-2}2^{1}}{1!})=54.13$$

$$e_2 = N\,p(X=2)=200(\frac{e^{-2}2^{2}}{2!})=54.13$$

$$e_3 = N\,p(X=3)=200(\frac{e^{-2}2^{3}}{3!})=36.09$$

$$e_4 = N\,p(X=4)=200(\frac{e^{-2}2^{4}}{4!})=18.04$$

$$e_5 = N\,p(X=5)=200(\frac{e^{-2}2^{5}}{5!})=7.22$$

$e_6 = N\,p(X=6)=200(\frac{e^{-2}2^{6}}{6!})=2.41$ كان من المفروض أن التكرارات لا تحتوي على كسور طالما أن وحدة القياس هي السفينة، غير انه يسمح بالكسور بالنسبة للتكرارات المتوقعة حتى لا تؤثر عملية التقريب على النتائج تأثيرا كبيرا وبذلك يكون مجموع التكرارات المتوقعة يساوي 199.09 سفينة ويكون الفارق الناتج هو ناتج التقريب.

ثانيا: اختبار حسن المطابقة:

(1) البيانات تتفق مع توزيع بواسون $H_0:$

البيانات لا تتفق مع توزيع بواسون $H_1:$

(2) حيث أن التكرار المتوقع عند $X = 6$ أقل من 5 فإننا ندمج هذه القيمة مع القيمة التي قبلها حتى نتمكن من تطبيق اختبار χ^2. وبذلك يمكن ترتيب القيم المشاهدة والمتوقعة في الجدول التالي.

X	0	1	2	3	4	≥ 5
التكرار المشاهد O_i	16	28	107	42	4	3
التكرار المتوقع e_i	27.07	54.13	54.13	36.09	18.04	9.63

وباستخدام جداول χ^2 نجد أن قيمة χ^2 هي: $\chi^2_{0.05,4} = 9.49$ حيث 4 هي درجات الحرية والتي تساوي:

$$d.f = r - 1 - k = 6 - 1 - 1 = 4$$

إذ تم طرح الواحد الأول لشرط المجموع الكلي للتكرارات ثم تم طرح الواحد الثاني نظرا لاعتمادنا على λ في إيجاد الاحتمالات وهو تقدير لمتوسط التوزيع في المجتمع تم استخراجه من بيانات العينة أساسا، وان $r = 6$ وهي عدد الأعمدة بعد عملية الضم.

(3) نحسب قيمة χ^2 - المحسوبة كما يلي:

$$\chi_c^2 = \sum_i \frac{(O_i - e_i)^2}{e_i}$$

$$= \frac{(16-27.07)^2}{27.07} + \frac{(28-54.13)^2}{54.13} + \cdots + \frac{(3-9.63)^2}{9.63} = 85.24$$

(4) ونظرا لأن χ^2 - المحسوبة أكبر من χ^2 - الجدولية، فإننا نرفض H_0 ونقبل H_1 وهذا يعني أن البيانات المتاحة لا تنتمي إلى توزيع بواسون وأن عملية التوفيق غير جيدة.

(ب) توفيق منحنى ذي الحدين: المثال التالي يوضح كيفية توفيق هذا المنحنى.

- **مثال (69)**: ألقيت 5 وحدات من النقود 80 مرة، وفي كل مرة كان يسجل عدد مرات ظهور الوجه وقد حصلنا على البيانات الآتية:

عدد مرات ظهور الوجه X	0	1	2	3	4	5
التكرار المشاهد O_i	2	4	5	16	30	23

والمطلوب:

(1) توفيق توزيع ذي الحدين.

(2) اختبار جودة التوفيق باحتمال 95%.

الحل:

<u>أولا</u>: توفيق توزيع ذي الحدين

من المعروف أن دالة توزيع ذي الحدين عند إلقاء 5 وحدات تأخذ الشكل التالي:

$$p(X = x) = \binom{5}{x} p^x q^{n-x}, \qquad x = 0,1,2,3,4,5$$

حيث p, q هو احتمال النجاح والفشل على التوالي. ولتحديد قيمة p من العينة فانه يمكن عمل ذلك باستخدام طريقة العزوم والتي مفادها إننا يمكن أن نقدر المعلمة p بمساواة العزم النظري المحسوب من الدالة بالعزم المحسوب من العينة كالآتي:

العزم المحسوب من الدالة = متوسط الدالة = $np = 5p$.

أما العزم المحسوب من العينة هو $\overline{X}$ ، حيث:

$$\overline{X} = \frac{\sum Xf}{\sum f} = \frac{0(2) + 1(4) + 2(5) + \cdots + 5(23)}{80}$$
$$= \frac{297}{80}$$

وبمساواة العزم المحسوب من العينة بالعزم المحسوب من الدالة نجد أن: $5p = \frac{297}{80}$،

ومنها نجد أن $p = 0.74$. وعليه فان دالة التوزيع الموفق لذي الحدين هي:

$$p(X = x) = \binom{5}{x}(0.74)^x (0.26)^{5-x}, \qquad x = 0,1,2,3,4,5$$

وباستخدام هذه الدالة يكون لدينا التكرارات المتوقعة الآتية.

X	0	1	2	3	4	5
التكرار المتوقع e_i	0.09	1.30	7.54	21.71	31.30	18.06

<u>ثانيا:</u> اختبار حسن المطابقة

لأن التكرارات المتوقعة للخلية ألأولى والثانية أقل من 5 فانه يتعذر استخدام χ^2 لإجراء الاختبار. وبناء عليه فانه يمكن أن نجري تعديلا في البيانات أو نبحث عن اختبار آخر. ولإجراء التعديل فلا شيء يتأثر سوى درجات الحرية فقط. فمثلا يمكن أن نضم التكرار المشاهد والمتوقع عند ما تكون $X = 0, 1, 2$ ثم نجري الاختبار كالآتي:

(1) البيانات تتبع توزيع ذي الحدين $H_0:$

البيانات لا تتبع توزيع ذي الحدين $H_1:$

(2) لإيجاد قيمة χ^2 - الجدولية يجب أولا أن نحسب χ^2 بعد عملية الدمج أو الضم لنرى التأثير الذي سيحدث على درجات الحرية كما يلي:

$\chi^2_{0.05,4-1-1} = \chi^2_{0.05,2} = 5.99$.

(3) نضع جميع القيم في جدول كالجدول أدناه ومنه نحسب قيمة χ^2 - المحسوبة التي تعبر عن مقدار الفرق بين التكرارات المشاهدة والتكرارات المتوقعة.

جدول (رقم 26)

X	التكرارات المشاهدة O_i	التكرارات المتوقعة e_i	$O_i - e_i$	$\frac{(O_i - e_i)^2}{e_i}$
2≥	11	8.93	2.07	0.48
3	16	21.71	- 5.71	1.50
4	30	31.30	- 1.30	0.05
5	23	18.06	4.94	1.35
المجموع	**80**	**80**		**3.38**

(4) وحيث أن χ^2 - الجدولية أكبر من نظيرتها المحسوبة نرفض الفرض البديل H_1 ونقبل الفرض العدم H_0 القائل بأنه لا يوجد اختلاف بين التوزيع النظري والتوزيع المشاهد وكليهما يتبع توزيع ذي الحدين.

(5.5) اختبار كولومجروف (*The Kolomgrov Test*)

ذكرنا إننا لا نستطيع إجراء اختبار χ^2 إلا بعد تحقيق الشرطين المذكورين سابقا، لذلك فان الاختبار الجديد المذكور لا يعتمد أساسا على الفرضين السابقين كما انه أكثر قوة من اختبار χ^2 حتى وان كانت الشروط مستوفية. ويعتمد هذا الاختبار على التوزيع الاحتمالي التجميعي لكل من التوزيع المتوقع والمشاهد. فإذا كان التوزيعان

يقتربان من بعضهما البعض فانه يمكن الجزم بقبول فرض العدم القائل بأن التوزيع النظري للمتغير يتفق مع التوزيع المشاهد. وتتوقف عمليات الرفض والقبول على الفرق بين التوزيعين، فان كان الفرق كبيرا كان فرقا معنويا والعكس صحيح إذا كان الفرق بين التوزيعين فرقا صغيرا . **ولإجراء الاختبار نقوم بعمل الخطوات التالية:**

(1) يحدد الفرض العدم والفرض البديل كالتالي:

$H_0:$

دالة الاحتمال التجميعي للتوزيع النظري (d_1) تساوي دالة الاحتمال التجميعي للتوزيع المشاهد (d_2)

$H_1:$

الدالتان مختلفتان

(2) نوجد الفرق المطلق d بين دالة الاحتمال التجميعي للتكرار المشاهد ودالة الاحتمال التجميعي للتكرار المتوقع أي: $d = |d_1 - d_2|$، وقد وجد أن الحد الأقصى لهذا الفرق كمتغير يتبع توزيعا معينا يطلق عليه توزيع d.

(3) هناك جداول خاصة (جداول كولمجروف) لاستخراج القيم الحرجة لهذا المتغير وتعتمد على عدد المفردات (حجم العينة) وعلى الاحتمال. وهذا معناه انه باحتمال معين يمكن استخراج قيمة جدولية للمتغير d.

(4) يقارن الحد الأقصى للفرق المطلق بين المشاهد والمتوقع بقيمة d المستخرجة من الجدول **ويكون لدينا الحالتين الآتيتين:**

1. إذا كان الفرق المحسوب أصغرمن d المستخرجة من الجدول يقبل الفرض العدم.
2. إذا كان الفرق المحسوب أكبرمن d المستخرجة من الجدول نرفض الفرض العدم.

مثال (70): البيانات الآتية هي نتائج إلقاء 12 زهرة من زهرات النرد .

المتغير X	1	2	3	4	5	6
التكرار المشاهد	0	1	1	1	5	4

والمطلوب اختبار هل هذه النتائج تتفق مع التوزيع المنتظم باحتمال 95%.

الحل:

أولا: توفيق منحنى منتظم

لتوفيق منحنى منتظم فإننا نفترض نظريا أن عدد مرات ظهور كل وجه متساوي وبذلك يكون لدينا التكرارات المتوقعة الآتية:

المتغير X	1	2	3	4	5	6
التكرار المشاهد	2	2	2	2	2	2

ثانيا: اختبار حسن المطابقة

لا يمكننا استخدام χ^2 لاختبار جودة التوفيق وذلك لأن مجموع التكرارات أقل من 30 أولا، ولأن التكرار المتوقع في كل خلية من الخلايا أقل من 5، ولذلك نقوم بإجراء اختبار كولومجروف والذي يوضحه جدول (رقم 26) خطوات حسابه.

(1)
$$H_0 : d_1 = d_2$$
$$H_1 : d_1 \neq d_2$$

(2) بحجم عينة 12 وباحتمال 5%، نوجد القيمة المقابلة لها من جدول كولومجروف كما يلي:

$$d_{\alpha,n} = d_{0.05,12} = 0.375$$

(3) من جدول (رقم 26) نجد أن الحد الأقصى للفرق المطلق بين دالة الاحتمال التجميعي للتكرار المشاهد والمتوقع هو:

$$d = |d_1 - d_2| = \frac{5}{12} = 0.417$$

(4) وحيث أن الاختبار من طرف واحد وأن d – المحسوبة أكبر من d – المستخرجة من الجدول نرفض الفرض العدم ونستنتج بأن هذه البيانات لا تتبع التوزيع المنتظم أو أن الزهرة غير متوازنة. وبنفس هذا الأسلوب يمكن استخدام هذا الاختبار في جميع اختبارات جودة التوفيق.

جدول (رقم 26): اختبار كولومجروف

X	التكرار المشاهد O_i	التكرار المتوقع e_i	الاحتمال المتجمع المشاهد d_1	الاحتمال المتجمع المتوقع d_2	الفرق المطلق بين الاحتمالين $\|d_1 - d_2\|$
1	0	2	0	2/12	2/12
2	1	2	1/12	4/12	3/12
3	1	2	2/12	6/12	4/12
4	1	2	3/12	8/12	5/12
5	5	2	8/12	10/12	2/12
6	4	2	12/12	12/12	0
المجموع	**12**	**12**			

تمـــارين

(1) الجدول التالي يبين التوزيع التكراري لعدد الحوادث التي وقعت لـ 1000 عامل في أحد المصانع خلال سنة معينة.

عدد الحوادث	0	1	2	3	4	5
التكرار المشاهد	350	360	190	70	20	10

فهل ترى أن عدد الحوادث يخضع لتوزيع بواسون عند $\alpha = 5\%$ ؟

(2) أرسل استبيان إلى خريجي كلية الطب في جامعة البصرة. والجدول التالي يوضح أعداد الخريجين الذين أعادوا هذا الاستبيان:

الدرجة العلمية / الحالة	بكالوريوس	ماجستير	دكتوراه
أعادوا الاستبيان	45	40	44
لم يعيدوا الاستبيان	11	11	6

والمطلوب اختبار الفرضية باستقلالية العلاقة بين الدرجة العلمية وحالة إعادة الاستبيان عند $\alpha = 0.05$.

(3) استخدم نوعان من المبيدات لعلاج مرض فطري في بذور الشعير، فكانت النتائج كما يلي:

حالة البذور / نوع المبيد	بذور حية	بذور ميتة
ألأول	130	35
الثاني	90	60

فهل ترى أن هناك علاقة بين نوع المبيد وحالة البذور عند $\alpha = 0.05$.

(4) رمي حجر نرد 100 مرة فكانت النتائج كما في الجدول التالي:

الناتج	1	2	3	4	5	6
التكرار	12	21	15	12	19	21

هل تتفق مع الرأي القائل أن هذا الحجر غير منتظم عند $\alpha = 5\%$ ؟

(5) ألقيت 4 وحدات من النقود مائة مرة وكانت لدينا النتائج التالية:

عدد مرات الظهور	0	1	2	3	4
التكرار	17	23	20	24	16

المطلوب توفيق منحنى ذي الحدين واختبار جودة التوفيق باحتمال 99%.

(6) ألقيت عملة 10 مرات وكانت النتائج الآتية:

ظهر	وجه	**النتيجة**
6	4	**التكرار**

وباحتمال 99% يطلب منك أن ترى ما إذا كانت هذه العملة متزنة أم لا ؟

(7) تناول 150 رجل دواء مضادا للملاريا فأظهر 15 فردا حساسية له، كما أظهرت 40 امرأة من 140 امرأة تناولن نفس الدواء الأعراض ذاتها. حلل هذه النتائج لتحديد ما إذا كانت هناك علاقة بين الجنس البشري والحساسية من الدواء باحتمال قدره 0.975.

الفصل السادس

تحليل التباين

(1.6) مقدمة.

(2.6) التصنيف الأحادي.

(3.6) المقارنات المتعددة.

(4.6) تحليل التباين الثنائي.

تحليل التباين

Analysis of Variance

(1.6) مقدمة (*Introduction*)

تعرفنا في الباب الرابع على الاختبارات التي تناقش الفرق بين عينتين ولم نناقش مشكلة اختبارات الفروض لأكثر من عينتين، حيث أن هناك تجارب إحصائية كثيرة تتعلق باختبار تساوي معدلات ثلاثة مجتمعات أو أكثر. فلو فرضنا وجود 5 عينات وأريد اختبار الفرق بين هذه العينات باستخدام الأساليب السابق مناقشتها فإننا سوف نختبر هذه العينات زوجا زوجا، أي أننا سوف نجري 10 اختبارات $\binom{5}{2}=10$، الأمر الذي يتسم بالصعوبة وخصوصا لو زاد عدد المعالجات عن هذا العدد حيث نتكلف جهدا ووقتا وتكاليفا في سبيل إجراء هذه الاختبارات. كما أننا سوف نفشل في الوصول إلى حكم عام نستطيع عن طريقه أن نحدد أي من هذه العينات أحسن لأن كل اختبار سوف يكون له الاستنتاج الخاص به، فليس شرطا أن تكون كل الاختبارات معنوية أو غير معنوية، بالإضافة إلى ذلك كله فإننا إذا ما استعملنا مستوى الدلالة α يساوي 5% مثلا، فإن احتمال الوصول إلى القرار الصحيح $(1-\alpha)$ بعدم وجود فرق ذي دلالة لكل اختبار هو 95%، وبالتالي فإن احتمال الوصول إلى القرار الصحيح للاختبارات الخمسة هو $(0.95)^{10}$ مما يعني أن احتمال الوصول إلى قرار واحد غير سليم أو أكثر هو:

$$=1-(0.95)^{10}$$
$$=1-0.59=0.41$$

بمعنى أننا لو استعملنا اختبار t مثلا لكل زوج من المجتمعات الخمسة على حدة فإن نسبة اتخاذ قرار غير صحيح أو أكثر هو 41%، ولذلك كان من الضروري التفكير في طرق غير اختبار t في حالات كهذه. ومن هنا برزت فكرة اختبار تحليل التباين والذي يعتبر من أفضل الاختبارات الإحصائية وأحسنها حيث يقوم بإجراء هذه الاختبارات معا مما يقلل الوقت والجهد والتكاليف.

ولقد سمي هذا الاختبار بتحليل التباين لأنه مبني أساسا على دراسة الاختلاف والتباين بين وداخل المجموعات. أما التوزيع المستخدم في هذا الاختبار فهو توزيع F، حيث يمكن القول أن من أهم تطبيقات توزيع F هو استخدامه في دراسة العلاقة بين عدة مجموعات ومعرفة ما إذا كان يوجد اختلاف حقيقي بين تلك المجموعات، أو أن هذا الاختلاف قد يرجع إلى عوامل عشوائية. ويمكن أن نسرد مثالا على ذلك كالآتي:

افرض أن لدينا 5 أنواع من الفيتامينات (أي 5 معالجات) من المفروض أنها تختلف عن بعضها في تأثيرها على زيادة الوزن. فإذا تم لنا أخذ 5 مجموعات مستقلة من الفئران مثلا تحت ظروف وشروط متشابهة تماما حتى يمكن إرجاع الاختلافات في الأوزان لاستخدام المعالجات المتاحة، فإننا في هذه الحالة نجري اختبار F لمعرفة مدى وجود اختلافات جوهرية بين متوسطات أوزان المجموعات الخمسة نتيجة أخذ المعالجات الجديدة (كل مجموعة معالجة بالطبع)، أم أن استخدام هذه المعالجات لم يكن له أي تأثير معنوي على وجود اختلافات بين متوسطات أوزان هذه المجموعات الخمسة؟

هنا نجد نوعية واحدة من المعالجات (الفيتامينات) ويراد معرفة تأثير هذه النوعية من المعالجات على المجموعات تحت الاختبار. وهنا نقول أن لدينا عاملا واحدا (*Factor*) وله عدة مستويات (*Levels*) يراد توزيعها على متغير معين هو Y مثلا. وإذا ما صنفت البيانات بهذا الشكل قيل للتصنيف بأنه تصنيف أحادي.

(2.6) التصنيف الأحادي (*One- Way Classification*)

وهو إجراء اختبار F لتحليل التباين في حالة وجود عامل واحد فقط، ويتركز هذا الاختبار لدراسة العلاقة بين عدة مجموعات على فروض لا يمكن استخدامه بدونها. ومن أهم هذه الفروض التي سنفترض تحققها جميعا في هذا الباب هي:

(1) المجتمعات التي تسحب منها العينات لها تباينات غير معروفة غير أنها متجانسة، بمعنى أن المجتمعات لها نفس التباين.

(2) المجتمعات التي تسحب منها العينات تتوزع توزيعا معتدلا أو قريبا من الاعتدال بدرجة كافية.

(3) استقلال العينات المسحوبة تماما من هذه المجتمعات المعتدلة.

ولأن اختبار تحليل التباين كما ذكرنا يقوم على دراسة التباين بين وداخل المجموعات، لذلك يتحتم علينا إيجاد تقديرين مستقلين غير متحيزين لهذا التباين لكي نتمكن من إجراء الاختبار. **ويمكننا الحصول على هذين التقديرين للتباين كما يلي:**

(1) بناء على الافتراض الأول فإن $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$، وبالتالي يمكن إيجاد تقدير للمعلمة σ^2 من خلال تباينات العينات المسحوبة عشوائيا من هذه المجتمعات والمرجحة بدرجات الحرية، أي أن:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \cdots + (n_k - 1)S_k^2}{N - k}$$

حيث:

$S^2 =$ تقدير غير متحيز للمعلمة σ^2.

$k =$ عدد العينات (المجموعات) المسحوبة عشوائيا من المجتمعات الأصلية.

$S_i^2 =$ هو تباين كل عينة من العينات $i = 1, 2, \ldots, k$.

$N =$ جميع المفردات المكونة للتجربة وهو حاصل جمع أحجام كل العينات، أي أن:

$$N = n_1 + n_2 + \cdots + n_k$$

ولأن:
$$S_i^2 = \sum_{j=1}^{n_{i i}} (y_{ij} - \bar{y}_i)^2 / n_i - 1$$

فإن:
$$(n_i - 1)S_i^2 = \sum (y_{ij} - \bar{y}_i)^2$$

وبالتالي:
$$S^2 = \frac{\sum_{j=1}^{n_1}(Y_{1j} - \bar{Y}_1)^2 + \sum_{j=1}^{n_2}(Y_{2j} - \bar{Y}_2)^2 + \cdots + \sum_{j=1}^{n_k}(Y_{kj} - \bar{Y}_k)^2}{N - k}$$

$$S^2 = \sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2 / N - k$$

ويطلق على هذا التباين بالتباين داخل المجموعات (*Variation Samples Within the*) أو التباين العام وسوف نعطي له الرمز (S_W^2)، وهو التقدير الأول للمعلمة σ^2.

(2) التقدير الثاني للمعلمة σ^2 يمكن إيجاده بأسلوب مختلف وذلك باستخدام التباينات بين متوسطات العينات أي أن:

$$S^2 = \frac{\sum_{i=1}^{k} n_j (\bar{y}_i - \bar{y})^2}{k - 1}$$

حيث:

$\bar{y}_i =$ متوسط كل عينة من العينات المشمولة بالدراسة والتي عددها k.

$\bar{y} =$ المتوسط العام المحسوب من كل مفردات العينات.

$n_j =$ حجم كل عينة مع مراعاة أنها قد تختلف من مجموعة إلى أخرى.

ويطلق على هذا التباين بالتباين بين المجموعات (*Variation Between the Samples*) ونعطي له الرمز (S_B^2). هذا التقدير للمعلمة σ^2 يتأثر بالاختلافات بين

المتوسطات نتيجة لاستخدام معالجة معينة لأكثر من مستوى لها. وبالتالي فان قيمة S_B^2 تختلف عن المقدر S_W^2

كلما زادت الاختلافات بين المتوسطات الحسابية في المجتمعات ألأصلية. وأي اختلاف بين S_B^2 & S_W^2 إنما يرجع أساسا إلى عدم صحة الفرض القائل بأن العينات قد سحبت من مجتمعات لها أوساط حسابية متساوية مما أدى إلى اختلاف متوسطات العينات، وبالتالي فان المقارنة بين هذين التقديرين بإيجاد النسبة بينهما تعتبر أساسا لاختبار F ، حيث أن:

$$F_c = \frac{S_B^2}{S_W^2}$$

وكلما زادت هذه النسبة عن الواحد الصحيح كلما كان ذلك دليلا على وجود اختلافات كبيرة بين المتوسطات الحسابية التي سحبت منها العينات، لأن زيادة البسط عن المقام إنما ترجع حتما إلى وجود الاختلافات (المعنوية أو غير المعنوية) بين متوسطات المجتمعات التي تم سحب العينات منها عشوائيا.

وبالإضافة إلى الطريقة السابقة هناك طريقة ثانية لإجراء هذا الاختبار وذلك عن طريق إيجاد تقدير ثالث غير متحيز للمعلمة σ^2، وسوف نعطي له الرمز S^2 ونعرفه بالمعادلة التالية:

$$S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2}{N-1}$$

والآن إذا أخذنا البسط في كل من التقديرات الثلاثة للمعلمة σ^2 (أي البسط من تقدير S_W^2, S_β^2, S^2) فإننا نصل إلى العلاقة الآتية:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2 = \sum_{i=1}^{k} n_j(\bar{Y}_i - \bar{Y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2 \qquad (1)$$

حيث $\bar{Y}$ هو المتوسط العام ويعرف كما يلي:

$$\bar{Y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}}{N}$$

وأن $\bar{Y}_i$ هو متوسط المعالجة i، أي أن:

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_j}$$

ويمكن التعبير عن مجموعات المربعات في العلاقة (1) باستعمال الرموز الآتية:

(1) المجموع الكلي للمربعات: *Total Sum of Square (SST)*

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2$$
$$= \sum_{i}\sum_{j} Y_{ij}^2 - N\bar{Y}^2$$

حيث $N\bar{Y}^2$ هو معامل التصحيح *Correction Factor (C.F)* والذي يمكن أن يعرف كذلك كما يلي:

$$C.F = \frac{(\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij})^2}{N}$$

(2) مجموع مربعات المعالجات: *Sum of Square for treatments (SSt)*

$$SSt = \sum_{i=1}^{k} n_j\,(\bar{Y}_i - \bar{Y})^2$$
$$= \sum_{i=1}^{k} n_j\,\bar{Y}_i^{\,2} - N\bar{Y}^{\,2}$$

(3) مجموع مربعات الخطأ: *Sum of Square for Error (SSE)*

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2$$

أو أن:

$$SSE = SST - SSt$$

ومن هذا كله نجد أن أول تقدير مبني على (*k-1*) من درجات الحرية للتباين σ^2 هو S_1^2 حيث:

$$S_1^2 = \frac{SSt}{k-1} = S_\beta^2$$

وهو تقدير غير متحيز لـ σ^2 إذا كان فرض العدم صحيحا. أما إذا كان H_0 غير صحيح فإن *SSt* سوف تكون كبيرة وسوف تقدر التباين σ^2 بقيمة أعلى مما يجب. أما التقدير الثاني للتباين σ^2 فهو S_e^2 حيث:

$$S_e^2 = \frac{SSE}{N-k} = S_W^2$$

وهو تقدير مستقل عن التقدير الأول ويقوم على *N - k* من درجات الحرية حيث يعتبر تقديرا غير متحيز للمعلمة σ^2 عندما تكون H_0 صحيحة أو غير صحيحة.

وأخيرا فإن تباين البيانات المجمعة هو:

$$S^2 = \frac{SST}{N-1}$$

وهو تقدير غير منحاز لـ σ^2 أيضا عندما تكون H_0 صحيحة، وهو مبني على *N-1* من درجات الحرية. وبشكل عام نلاحظ أن درجات الحرية تقسم إلى قسمين مرتبطين بكل مجموع من مجاميع المربعات في العلاقة (1) كما يلي:

$$N - 1 = (k - 1) + (N - k)$$

وبناء على ذلك كله وعندما تكون H_0 صحيحة فإن المتغير العشوائي F حيث:

$$F_c = \frac{S_1^2}{S_e^2} = \frac{SSt/(k-1)}{SSE/(N-k)} \qquad (2)$$

يخضع لتوزيع F بدرجات حرية $(k - 1)$ و $(N - k)$. ولأن S_1^2 تميل لتقدير المعلمة σ^2 بقيمة أعلى مما يجب عندما تكون H_1 صحيحة، لذلك فنحن نستعمل الطرف الأيمن فقط من توزيع F كمنطقة حرجة لاختبار H_0. وبالتالي نرفض الفرض العدم H_0 على مستوى المعنوية α إذا كان:

$$F_c \geq f_{\alpha,(k-1)(N-k)}$$

لاحظ أن العلاقة (2) تكافئ الآتي:

$$F_c = \frac{R^2/k-1}{(1-R^2)/N-k}$$

حيث R^2 هو معامل التحديد الذي تعرفه المعادلة التالية:

$$R^2 = \frac{SSt}{SST} = 1 - \frac{SSE}{SST}$$

ويمكن تصوير جدول تحليل التباين الأحادي كما يلي.

جدول (رقم 27)

جدول تحليل التباين الأحادي (*One-Way ANOVA Table*)

مصادر التغير S.O.V	مجموع المربعات SS	درجات الحرية df	معدلات المربعات MS	F المحسوبة
المعالجات (بين المتوسطات)	SSt	k-1	$MSt = S_1^2 = \frac{SSt}{k-1}$	$F_c = \frac{S_1^2}{S_e^2}$
الخطأ (داخل المجموعات)	SSE	N-k	$MSE = S_e^2 = \frac{SSE}{N-k}$	
التغير الكلي	SST	N-1		

وضع الفروض لاختبارات F للعلاقة بين عدة مجموعات

ذكرنا فيما سبق أنه من المستحسن إجراء اختبار F على النحو المشروع سابقا في حالة وجود أكثر من مجموعتين ويراد اختبار العلاقة بينهما. ومن خلال تطبيق هذا الاختبار يمكن معرفة إلى أي مدى يوجد اختلاف بين متوسطات المجتمعات الحقيقية والتي تم سحب العينات منها. فإن وجد

اختلاف حقيقي بين متوسطاتها يقال أن العينات لا يمكن اعتبارها مسحوبة من مجتمعات لها متوسطات متساوية. أما إذا عرفنا من خلال إجراء الاختبار عدم وجود اختلاف حقيقي بين هذه المتوسطات فسوف نقبل الفرضية القائلة بأن هذه العينات قد تم سحبها فعلا من مجتمعات لها متوسطات حسابية غير مختلفة عن بعضها وعلى ذلك فإن الفروض التي يمكن وضعها هنا تأخذ الشكل التالي:

(1) فرض العدم: ويكون كالآتي:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

حيث k هي عدد المجموعات أو المجتمعات الإحصائية تحت الاختبار.

(2) فرض البديل: ويكون كما يلي:

$$H_1 : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$$

وهنا يمكن قبول H_0 ورفض H_1 إذا كانت $F-$ المحسوبة أصغر من $f-$ الجدولية أو يقبل H_0 عندما:

$$F_c < f_{\alpha,(k-1,N-k)}$$

كما أننا نرفض H_0 وليس هناك داعي من قبوله عندما تكون:

$$F_c \geq f_{\alpha,(k-1,N-k)}$$

- **مثال (71)**: لدراسة أثر ثلاث طرق للمطالعة على القدرة في التحصيل، تم اختبار ثلاث مجموعات من الطلبة في إحدى الكليات (تحت ظروف معينة واحدة) كل مجموعة تتكون من 5 طلاب ثم سجلت المشاهدات أدناه لدرجاتهم من امتحان في مادة الإحصاء. المطلوب اختبار مدى وجود فرق جوهري لاستخدامهم هذه الأساليب الثلاث على درجة امتحان مادة الإحصاء على مستوى المعنوية 5%.

III المطالعة ثلاث ساعات نهارا وأربع ساعات ليلا	II المطالعة أربع ساعات نهارا وثلاث ساعات ليلا	I المطالعة سبعة ساعات في أي وقت
93	85	58
97	95	59
96	96	61
86	83	64
94	75	70

الحل:

نلاحظ في مثالنا أعلاه تساوي حجوم العينات المختلفة أي أن:

$$n_1 = n_2 = n_3 = n$$

(1) الفرض العدم H_0 : لا يوجد فرق جوهري بين المتوسطات الحسابية للمجتمعات التي سحبت منها هذه العينات، أو أن: $H_0 : \mu_1 = \mu_2 = \mu_3$.

الفرض البديل H_1 : يوجد فرق جوهري بين المتوسطات الحسابية للمجتمعات التي سحبت منها هذه العينات، أو أن: $H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

بمعنى أن تأثير الأساليب الثلاثة في المطالعة مختلف ولا يؤدي إلى نفس النتائج.

(2) ولإجراء الاختبار تحسب البيانات أدناه والمثبتة في جدول (رقم 28):

(أ) التباين داخل المجموعات S_w^2 يساوي:

$$S_w^2 = \frac{\sum_1^5 (Y_{1j} - \bar{Y}_1)^2 + \sum_1^5 (Y_{2j} - \bar{Y}_2)^2 + \sum_1^5 (Y_{3j} - \bar{Y}_3)^2}{15-3}$$

$$= \frac{93.2 + 308.8 + 74.8}{12} = \frac{476.8}{12} = 39.73$$

بدرجات حرية = عدد المشاهدات – عدد المعالجات

$= N - k$

$= 15\text{-}3$

$= 12$

(ب) التباين بين المجموعات S_β^2 يساوي:

$$S_\beta^2 = \frac{n \sum_{i=1}^{3} (\bar{Y}_i - \bar{Y})^2}{3-1}$$

$$= \frac{n\left[(\bar{Y}_1 - \bar{Y})^2 + (\bar{Y}_2 - \bar{Y})^2 + (\bar{Y}_3 - \bar{Y})^2\right]}{2}$$

$$= \frac{5\left[(62.4 - 80.8)^2 + (86.8 - 80.8)^2 + (93.2 - 80.8)^2\right]}{2} = \frac{2641.6}{2} = 1320.8$$

بدرجات حرية = عدد المعالجات – 1

$= k - 1$

$=3 - 1$

$=1$

جدول (رقم 28)

الأسلوب الأول I	الأسلوب الثاني II	الأسلوب الثالث III	مجموع المشاهدات $N = n_1 + n_2 + n_3$
58	85	93	58
59	95	97	59
61	96	96	61
64	83	86	64
70	75	94	70
			85
			95
			96
			83
			75
			93
			97
			96
			86
			94
$\sum Y_1 = 312$	$\sum Y_2 = 434$	$\sum Y_3 = 466$	$\sum_i \sum_j Y_{ij} = 1212$
$\bar{Y}_1 = \sum Y_1 / n = 312/5 = 62.4$	$\bar{Y}_2 = \sum Y_2 / n = 434/5 = 86.8$	$\bar{Y}_3 = \sum Y_3 / n = 466/5 = 93.2$	$\bar{Y} = \sum_i \sum_j Y_{ij} / N = 1212/15 = 80.8$
$\sum_{j=1}^{n} (Y_{1j} - \bar{Y}_1)^2 = 93.2$	$\sum_{j=1}^{n} (Y_{2j} - \bar{Y}_2)^2 = 308.8$	$\sum_{j=1}^{n} (Y_{3j} - \bar{Y}_3)^2 = 74.8$	

ثم نوجد قيمة F - المحسوبة كالآتي:

$$F_c = \frac{S_B^2}{S_W^2} = \frac{1320.8}{39.73} = 33.24$$

(3) يمكن إيجاد قيمة f من الجدول كما يلي:

$$f_{\alpha,(k-1)(N-k)} = f_{0.05,(2,12)} = 3.88$$

(4) نظرا لأن F - المحسوبة أكبر من f - الجدولية، نرفض الفرض العدم ونقبل الفرض البديل القائل بوجود اختلاف حقيقي في تأثير الأساليب الثلاثة للمطالعة على درجة امتحان الإحصاء. ويمكن تلخيص هذه المراحل في جدول تحليل التباين التالي.

جدول (رقم 29)

جدول تحليل التباين الأحادي (حالة تساوي حجوم العينات)

S.O.V	*SS*	*df*	*MS*	F_c
Between	2641.6	3-1=2	1320.8	33.24
Within	476.8	15-3=12	39.73	
Total	3118.4	15-1=14		

- **مثال (72)**: أعطيت كل مجموعة من المجموعات الخمسة من الفئران نوعا معينا من الفيتامين وقيست الزيادة في الوزن في كل مجموعة، فكانت النتائج الآتية:

فيتامين *E*	فيتامين *D*	فيتامين *C*	فيتام ين *B*	فيتامين *A*
10	0	16	12	2
10	3	13	10	10
	0	12	14	8
	5	10		0
		9		

والمطلوب باحتمال 95% اختبار ما إذا كان هناك فرق معنوي بين الفيتامينات المختلفة .

الحل: لاحظ عدم تساوي حجوم العينات المأخوذة على المعالجات المختلفة في هذا المثال، أي أن:

$$N = n_1 + n_2 + n_3 + n_4 + n_5$$
$$= 4 + 3 + 5 + 4 + 2 = 18$$

ولإعداد جدول تحليل التباين نستخدم الطريقة الثانية لذلك الغرض كما هو آت.

(1) يأخذ الفرض العدم والفرض البديل الأشكال الآتية:

$$H_0 : \mu_a = \mu_b = \mu_c = \mu_d = \mu_e$$
$$H_1 : \mu_a \neq \mu_b \neq \mu_c \neq \mu_d \neq \mu_e$$

وهذا يعني أننا نفترض أن تأثير المعالجات الخمسة (الفيتامينات) متساوي أو يؤدي إلى نفس الزيادة في الوزن إذا كانت نتيجة الاختبار غير معنوية.

(2) لإجراء الاختبار تحسب البيانات التالية المطلوبة في جدول تحليل التباين الأحادي.

$$\bar{Y}_a = \frac{2+10+8+0}{4} = 5, \qquad \bar{Y}_d = \frac{0+3+0+5}{4} = 2$$
$$\bar{Y}_b = \frac{12+10+14}{3} = 12, \qquad \bar{Y}_e = \frac{10+10}{2} = 10$$
$$\bar{Y}_c = \frac{16+13+12+10+9}{5} = 12, \qquad \bar{Y} = \frac{20+36+60+8+20}{18} = 8$$

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}^2 - N\bar{Y}^2$$
$$= ((2)^2 + (10)^2 + \ldots + (10)^2) - 18(8)^2$$
$$= 1592 - 1152$$
$$= 440$$

بدرجات حرية = عدد المفردات – 1

= 18 -1 = 17

$$SSt = \sum_{i=1}^{k} n_j Y_i^2 - N\bar{Y}^2$$

$$= 4(5)^2 + 3(12)^2 + 5(12)^2 + 4(2)^2 + 2(10)^2 - 18(8)^2$$

$$= 1468 - 1152$$

$$= 316$$

بدرجات حرية = عدد المعالجات –1

= 5 - 1 = 4

$$SSE = SST - SSt$$

$$= 440 - 316$$

$$= 124$$

بدرجات حرية = درجة حرية الكلي – درجة حرية المعالجات

= 17 – 4 = 13

ويمكن تلخيص النتائج السابقة في جدول تحليل التباين الآتي:

جدول (رقم 30)

جدول تحليل التباين الأحادي (حالة عدم تساوي حجوم العينات)

مصدر التغير S.O.V	مجموع المربعات SS	درجات الحرية df	معدل المربعات MS	قيمة F المحسوبة
بسبب المعالجات بسبب الخطأ	316 124	4 13	79 9.54	$F_c = \frac{79}{9.54} = 8.28$
المجموع الكلي	440	17		

كما يمكننا إيجاد قيمة F– المحسوبة عن طريق حساب معامل التحديد R^2 كما يلي:

$$R^2 = \frac{SSt}{SST} = \frac{316}{440} = 0.72$$

$$F_c = \frac{R^2/(k-1)}{(1-R^2)/(N-k)}$$
$$= \frac{(0.72)/4}{(1-0.72)/13} = 8.36$$

وليس الفارق بين النتيجتين سوى فارق التقريب لا غير.

(3) وباستخدام قيمة f من الجدول فإننا على أساسها يمكننا رفض أو قبول فرض العدم، هذه القيمة هي:

$$f_{\alpha,(k-1,N-k)} = f_{0.05,(4,13)} = 3.14$$

(4) بمقارنة $F-$ المحسوبة من العينة بقيمة $f-$ الجدولية نجد أن الفرق معنوي، وبذلك نرفض فرض العدم ونقبل الفرض البديل أي أن هناك بعضا أو كل المعالجات لها تأثير معنوي. **ويلاحظ على هذا الاختبار في المثالين السابقين ما يأتي:**

(أ) يطلق على هذا الاختبار اختبار F نسبة إلى التوزيع العيني المستخدم ونسبة إلى مكتشفه الإحصائي *Fisher*.

(ب) جميع القيم الموجودة في جدول تحليل التباين هي قيم موجبة ولا يمكن أن تكون سالبة.

(ج) إذا كانت قيمة $F-$ المحسوبة أقل من واحد، فلا داعي للكشف في الجدول حيث يمكن الاستنتاج مباشرة أن الفرق غير معنوي.

(د) يمكن طرح مقدار ثابت من كل مفردة في الأمثلة السابقة أو يمكن إضافة أي مقدار ثابت لكل مفردة حيث لن تؤثر هذه العملية في النتيجة النهائية.

(هـ) يمكن ضرب (أو قسمة) أي رقم في الأمثلة السابقة في مقدار ثابت، ولكن يجب قسمة مجموع المربعات لكل من الكلي والمعالجات والخطأ على مربع ذلك الرقم. والعكس صحيح بالنسبة للقسمة (خواص التباين).

(3.6) المقارنات المتعددة (*Multiple Comparisons*)

مما لا شك فيه أن طريقة تحليل التباين طريقة قوية لاختبار تساوي أكثر من معدلين، غير أن هذا الاختبار إجمالي وهو يعجز عن تحديد أي المعالجات السبب في معنوية الاختبار عندما نرفض الفرض العدم ونقبل الفرض البديل. بمعنى آخر عندما يكون هناك تأثير معنوي من بعض أو كل المعالجات فإن الاختبار لا يحدد أي من معدلات المجتمعات متساوية وأي منها غير متساوية. ولقد تم إثبات أنه عندما يكون اختبار تحليل التباين معنويا فانه على الأقل يوجد اختبار واحد من اختبارات الفرق بين متوسطي عينتين (والتي يمكن إجراءها من كل الاختبارات الممكنة) تكون نتيجته معنوية.

وهناك اختبارات عدة يتطلب إجراء واحد منها كاختبار لاحق لاختبار تحليل التباين لتحديد أي من المعالجات السبب في معنوية اختبار تحليل التباين. ومن هذه الاختبارات:

(1) اختبار الحد الأدنى للفرق (الفرق المعنوي الأصغر) *L.S.D*

(2) اختبار t – المتعدد *Multiple t-test*

(3) اختبار شفييه *Scheffe's test*

<u>أولا:</u> اختبار الحد الأدنى للفرق

سوف نقوم بشرح هذا الاختبار وذلك بتطبيق خطوات الحل على مثال سابق.

- **مثال (73)**: بالعودة إلى مثال (72)، المطلوب التعرف على المعالجات التي تسببت في معنوية اختبار تحليل التباين باستخدام اختبار الحد الأدنى للفرق.

الحل: في مثال (72) كان لدينا 5 متوسطات لخمسة معالجات وهي:

$$\bar{Y}_a = 5, \quad \bar{Y}_b = 12, \quad \bar{Y}_c = 12, \quad \bar{Y}_d = 2, \quad \bar{Y}_e = 10$$

ومن الواضح أن هناك فروقا بين هذه المتوسطات، والسؤال هل هذه الفروق معنوية أم لا ؟ وللإجابة على هذا السؤال سوف نقوم بإجراء عشرة اختبارات: $\binom{5}{2}=10$ لمعرفة هل هناك فرق بين كل مجموعتين أم لا ؟ هذه الاختبارات هي:

$$\mu_a = \mu_b,\quad \mu_a = \mu_c,\quad \mu_a = \mu_d,\quad \mu_a = \mu_e.$$
$$\mu_b = \mu_c,\quad \mu_b = \mu_d,\quad \mu_b = \mu_e.$$
$$\mu_c = \mu_d,\quad \mu_c = \mu_e.$$
$$\mu_d = \mu_e.$$

وسنوضح الاختبار لبعض منها وسوف نترك الباقي كتمرين للقارئ. وبالتالي لكي نختبر الفرق بين متوسط المجموعة الأولى والثانية نقوم بالخطوات الآتية:

(1) نختبر الفرض العدم: $H_0 : \mu_a = \mu_b$

مقابل الفرض البديل: $H_1 : \mu_a \neq \mu_b$

(2) نقوم بحساب الحد الأدنى للفرق بين المجموعتين (A , B) حيث:

$$L.S.D = t_{\alpha/2,N-k}\sqrt{MSE(\frac{1}{n_a}+\frac{1}{n_b})}$$

حيث MSE تمثل متوسط مربعات الخطأ.

n_a تمثل حجم المجموعة الأولى.

n_b تمثل حجم المجموعة الثانية.

أما $t_{\alpha/2,N-k}$ فهي قيمة من الجدول بدرجات حرية الخطأ وهي تساوي:

$$t_{0.05/2,13} = 2.16$$

أي أن:

$$L.S.D = (2.16)\sqrt{9.54(\frac{1}{4}+\frac{1}{3})} = 5.09$$

(3) نقارن هذه القيمة بالفرق المطلق بين متوسط المجموعة الأولى والثانية، ويحسب هذا الفرق كالتالي:

$$\left|\overline{Y}_a - \overline{Y}_b\right| = \left|5-12\right| = 7$$

(4) إذا كانت قيمة الحد الأدنى للفرق أكبر من قيمة الفرق المطلق بين المجموعتين الأولى والثانية، فان الفرق بين المجموعتين فرق غير معنوي والعكس صحيح. ولأن الفرق المطلق يساوي 7 وهو أكبر من الحد الأدنى للفرق 5.09 فهذا يعني أن الفرق بين المتوسطين معنوي وبالتالي فإنهما مشتركان في معنوية F.

وبنفس الأسلوب نقوم بالاختبار بالنسبة للمجموعة الأولى والثالثة حيث:

(1) نختبر الفرض: $H_0 : \mu_a = \mu_c$ مقابل الفرض: $H_1 : \mu_a \neq \mu_c$

(2) الحد الأدنى للفرق بين المجموعتين (A, C) هو:

$$L.S.D = t_{\alpha/2, N-k}\sqrt{MSE(\frac{1}{n_a}+\frac{1}{n_c})}$$

$$= (2.16)\sqrt{9.54(\frac{1}{4}+\frac{1}{5})} = 4.48$$

(3) الفرق المطلق بين المتوسطين:

$$\left|\overline{Y}_a - \overline{Y}_c\right| = \left|5-12\right| = 7$$

(4) وبمقارنة قيمة الحد الأدنى للفرق بين المجموعتين بقيمة الفرق المطلق، نجد أن الفرق معنوي أيضا وان هذين المتوسطين مشتركان في معنوية F.

وهكذا نستمر في إجراء جميع الاختبارات اللازمة.

وكقاعدة عامة: الحد الأدنى للفرق ($L.S.D$) للمجموعتين (1, 2) يكون كالآتي:

$$L.S.D = t_{\alpha/2, N-k}\sqrt{MSE(\frac{1}{n_1}+\frac{1}{n_2})}$$

ملاحظة:

(أ) إذا كان اختبار F غير معنوي فلا داعي لإجراء اختبار الحد الأدنى للفرق، **ونستنتج** مباشرة أن الفرق غير معنوي بين كل المقارنات.

(ب) إذا كانت أحجام المعالجات متساوية (أي أن $n_1 = n_2 = \cdots = n_k = n$) فإننا نكتفي بحساب حد أدنى واحد للفرق ليستخدم في جميع المقارنات وهو في هذه الحالة يساوي:

$$L.S.D = t_{\alpha/2,N-k}\sqrt{MSE(\frac{2}{n})}$$

- **مثال (74):** بالعودة إلى مثال (71)، المطلوب استخدام اختبار الحد الأدنى للفرق للتعرف على المعالجات التي سببت معنوية اختبار تحليل التباين باحتمال 95%.

الحل: في مثال (71) كان لدينا 3 متوسطات لثلاثة معالجات وهي:

$$\bar{Y}_1 = 62.4, \quad \bar{Y}_2 = 86.8, \quad \bar{Y}_3 = 93.2$$

وسوف نقوم بإجراء 3 اختبارات $\binom{3}{1} = 3$ وهي:

(1) $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

(2) $H_0 : \mu_1 = \mu_3$
$H_1 : \mu_1 \neq \mu_3$

(3) $H_0 : \mu_2 = \mu_3$
$H_1 : \mu_2 \neq \mu_3$

ثم نقوم بحساب:

(1) الحد الأدنى للفرق بين المجموعات (وهو حد أدنى واحد لتساوي حجوم العينات):

$$L.S.D = t_{\alpha/2,N-k}\sqrt{MSE(\frac{2}{n})}$$

$$= t_{0.025,12}\sqrt{39.73(\frac{2}{5})} = 8.69$$

(2) الفرق المطلق بين متوسطات المجموعات الثلاثة:

$$(i) \quad |\bar{Y}_1 - \bar{Y}_2| = |62.4 - 86.8| = 24.4$$

$$(ii) \quad |\bar{Y}_1 - \bar{Y}_3| = |62.4 - 93.2| = 30.8$$

$$(iii) \quad |\bar{Y}_2 - \bar{Y}_3| = |86.8 - 93.2| = 6.4$$

(3) بمقارنة الحد الأدنى للفرق بين المجموعات (8.69) مع الفروق المطلقة لمتوسطات المجموعات نستنتج الآتي:

(أ) بالنسبة للمجموعتين الأولى والثانية نجد أن الفرق المطلق أكبر من الحد الأدنى للفرق مما يعني أن الفرق بين المتوسطين معنوي وبالتالي فإن المجموعة الأولى والثانية مشتركان في معنوية F.

(ب) بالنسبة للمجموعتين الأولى والثالثة نجد أن الفرق غير معنوي.

(ج) وأخيرا نجد أن الفرق معنوي بين المجموعتين الثانية والثالثة ذلك لأن الفرق المطلق بين المتوسطين أكبر من الحد الأدنى للفرق وبالتالي فهما مشتركان في معنوية F.

ثانيا: اختبار t– المتعدد

ويسمى اختبار بونفيروني، وهو يقوم على أساس إجراء اختبار t على الفرق بين كل معدلين على أن يكون مستوى الدلالة الكلي لجميع الاختبارات معا هو α . ففي مثال (72) وجدنا أننا نحتاج إلى ($r = 10$) من الاختبارات، وبالتالي فإننا نجري هذه الاختبارات على أن يكون مستوى المعنوية لكل اختبار هو:

$$\frac{\alpha}{r} = \frac{0.05}{10} = 0.005$$

وهذا الاختبار يصلح للاستعمال في حالة تساوي أو عدم تساوي حجوم العينات.

- **مثال (75):** في مثال (72) كانت نتيجة اختبار تحليل التباين هو رفض H_0. والمطلوب معرفة أي من هذه المعدلات كانت السبب في هذا الرفض على مستوى المعنوية 10% باستخدام اختبار t – المتعدد.

الحل:

(1) الفرضيات التي سوف نقوم باختبارها هي الفرضيات المذكورة في مثال (72).
وكما ذكرنا فإن $r = 10$ ومنها نجد أن مستوى المعنوية لكل اختبار من اختبارات t هو 0.01 حيث:

$$\frac{\alpha}{r} = \frac{0.1}{10} = 0.01$$

ولكي نقوم بإجراء الاختبار للفرضية الأولى:

$$H_0 : \mu_a = \mu_b$$
$$H_1 : \mu_a \neq \mu_b$$

فإننا نستعمل مستوى المعنوية 0.01 لإيجاد قيمة t من الجدول.

(2) لإيجاد قيمة t – المحسوبة نستعمل دالة الاختبار الآتية:

$$t_c = \frac{(\bar{Y}_a - \bar{Y}_b) - (\mu_a - \mu_b)}{\sqrt{MSE(\frac{1}{n_a} + \frac{1}{n_b})}}$$

وبالتالي فإن قيمة t هي:

$$t_c = \frac{(5-12)-0}{\sqrt{9.54(\frac{1}{4} + \frac{1}{3})}} = -2.967$$

(3) نقارن هذه القيمة بقيمة t – الجدولية وهي:

$$\pm t_{\alpha/2,N-k} = \pm t_{0.01/2,13} = \pm t_{0.005,13} = \pm 3.01$$

(4) ونظرا لأن قيمة t – المحسوبة (2.967-) أكبر من قيمة t – الجدولية (-3.01) نقبل H_0

ونستنتج أن $\mu_a = \mu_b$.

وهكذا بالنسبة لبقية المجموعات حيث يحسب المختبر الإحصائي t_c كما في الفقرة (2)، ثم نقارن هذه القيمة مع قيمة t – الجدولية (± 3.01) لاتخاذ القرار بصدد تساوي أو عدم تساوي المعدلات المتبقية.

ملاحظة:

يمكن إجراء اختبار t – المتعدد في حالة تساوي حجوم العينات، وكل ما في الأمر أننا نستعمل $MSE(\frac{2}{n})$ كتقدير لتباين $\bar{Y}_a - \bar{Y}_b$ بدلا من استعمال $MSE(\frac{1}{n_a}+\frac{1}{n_b})$.

ثالثا: اختبار شفييه *Scheffe's test*

وهو الاختبار الثالث الذي نستعمله لمعرفة تساوي أو عدم تساوي معدلات المجتمعات. فإذا كان لدينا k من المعالجات فإننا وتحت شروط وفرضيات التصنيف الأحادي نحتاج إلى إجراء المقارنات المتعددة بين كل زوج من المعدلات $\mu_1, \mu_2, \dots, \mu_k$ تحت مستوى معنوية كلي وهو α ، ثم نقوم بالخطوات الآتية لإجراء الاختبار.

(1) نضع الفرض العدم بالشكل التالي: $H_0 : \mu_i - \mu_j = 0$

مقابل الفرض البديل $H_1 : \mu_i - \mu_j \neq 0$

(2) نوجد قيمة f التي تعرف كالآتي:

$$f = \frac{(\bar{Y}_i - \bar{Y}_j)^2}{MSE(\frac{2}{n})}$$

إذا كانت حجوم العينات متساوية. أما إذا كانت حجوم العينات غير متساوية فان قيمة f تعرف كالآتي:

$$f = \frac{(\bar{Y}_i - \bar{Y}_j)^2}{MSE(\frac{1}{n_i} + \frac{1}{n_j})}$$

(3) من جدول F نوجد قيمة $f-$ الجدولية كما يلي: $f_{\alpha,(k-1)(N-k)}$.

ثم نضرب قيمة $f-$ الجدولية بالعدد $(k-1)$، حيث k عدد المعالجات. فإذا رمزنا للناتج w^2 فإن:

$$w^2 = (k-1) f_{\alpha,(k-1,N-k)}$$

(4) يرفض الفرض العدم H_0 إذا كانت قيمة f أكبر من قيمة w^2 والعكس صحيح. وهنا يمكن أن $\mu_i > \mu_j$ إذا كان:

$$\frac{\bar{Y}_i - \bar{Y}_j}{\sqrt{MSE(\frac{2}{n})}} > w$$

كما يمكننا اعتبار أن $\mu_i < \mu_j$ إذا كان:

$$\frac{\bar{Y}_i - \bar{Y}_j}{\sqrt{MSE(\frac{2}{n})}} < -w$$

حيث أن $w = \sqrt{w^2}$.

- **مثال (76):** عودة إلى مثال (71) حيث وجد أن نتيجة اختبار تحليل التباين معنوية. باستعمال اختبار شفييه، قرر أي من هذه المعدلات كانت سببا في هذه المعنوية على مستوى دلالة %5 ؟

الحل: لكي نقوم بإجراء الاختبار فإننا نقوم بالخطوات الآتية:

(1) نختبر الفرض الأول: $H_0 : \mu_1 - \mu_2 = 0$

مقابل الفرض الثاني: $H_1 : \mu_1 - \mu_2 \neq 0$

(2) نجد قيمة f وهي:

$$f=\frac{(\bar{Y}_1-\bar{Y}_2)^2}{MSE(\frac{2}{n})}=\frac{(62.4-86.8)^2}{39.73(\frac{2}{3})}=37.46$$

(3) نجد قيمة w^2 كما يلي:

$$w^2=(k-1)\,f_{\alpha,(k-1,N-k)}$$
$$=(3-1)\,f_{0.05,(2,12)}=2(3.88)=7.76$$

(4) لأن قيمة f أكبر من قيمة w^2، نرفض الفرض العدم H_0 ونستنتج أن المعدلين

μ_2 , μ_1

مختلفان، وهما يشتركان في معنوية F.

ولكي نعرف ما إذا كان μ_1 أصغر من μ_2 حيث $(\mu_1=62.4,\ \mu_2=86.8)$ فإننا نجد القيمة:

$$\frac{\bar{Y}_1-\bar{Y}_2}{\sqrt{MSE(\frac{2}{n})}}=\frac{62.4-86.8}{\sqrt{39.73(\frac{2}{5})}}=-6.12$$

ثم نقارن هذه القيمة بالقيمة:

$$-\sqrt{w^2}=-\sqrt{7.76}=-2.79$$

ولأن 2.79- > 6.12- نستنتج أن $\mu_1<\mu_2$.

وهكذا نكرر الخطوات السابقة إذا أردنا إجراء الاختبار الثاني والذي يكون كالتالي:

(1) نجري الاختبار للفرضية: $H_0:\mu_1-\mu_3=0$

مقابل الفرضية: $H_1:\mu_1-\mu_3\neq 0$

(2) نحسب f كما يلي:

$$f = \frac{(\bar{Y}_1 - \bar{Y}_3)^2}{MSE(\frac{2}{n})} = \frac{(62.4 - 93.2)^2}{39.73(\frac{2}{5})} = 59.69$$

(3) ولأن $w^2 = 7.76$، نجد أن $f > w^2$ في هذه الحالة أيضا، لذلك نرفض H_0 ونستنتج أن μ_1, μ_3 مشتركان في معنوية F.

ويمكن كذلك التأكد ما إذا كانت $\mu_1 < \mu_3$ بنفس الطريقة التي تمت في الاختبار الأول، أي أننا نقارن القيمة:

$$\frac{62.4 - 93.2}{\sqrt{39.73(\frac{2}{5})}}$$

بالقيمة $-\sqrt{w^2}$ لنقرر النتيجة.

وأخيرا وبحساب القيمة f لاختبار الفرضية الأخيرة $H_0 : \mu_2 = \mu_3$ نجد أنها تساوي 2.577، وبمقارنة هذه القيمة بقيمة w^2 نجد أننا لا نرفض الفرض العدم ونستنتج أن المعدلين متساويان ولا يشتركان في معنوية F.

(4.6) تحليل التباين الثنائي: التأثيرات الثابتة

Two- Way ANOVA Fixed Effects

إذا كنا نستخدم تحليل التباين الأحادي لدراسة أثر عامل واحد على متغير ما، فإننا يمكن أن نستخدمه في بعض الحالات التي قد يحتاج الباحث فيها إلى دراسة استخدام عاملين أو أكثر على متغير معين. كما لو أراد المهندس الزراعي معرفة تأثير عدة طرق للري وعدة أنواع من السماد المستخدم وعدة أنواع من المبيدات الحشرية على إنتاجية الأرض الزراعية. ويمكن استعمال طريقة تحليل التباين هذه في كل مجال كالاقتصاد والتربية والزراعة بالإضافة إلى الصناعة وعلم النفس وغيرها. وفي جميع

هذه الحالات يمكن استعمال تحليل التباين لدراسة تأثير مستويات كل عامل على حدة بالإضافة إلى دراسة تأثير هذه العوامل معا.

وفي هذا البند سنقوم بدراسة التجارب المؤلفة من عاملين فقط نسمي العامل الأول الصفوف ونسمي العامل الثاني الأعمدة، **مفترضين**:

(1) أن لكل عامل مستوى واحد فقط (أي أن عدد المشاهدات لكل خلية يساوي واحد فقط).

(2) أن العاملين (الأعمدة والصفوف) غير متفاعلين.

وفي هذه الحالة فإن مشاهدة من المشاهدات يمكن أن تكتب على النحو التالي:

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \qquad (3)$$

حيث:

α_i تعبر عن تأثير المجتمع i.

β_j تعبر عن تأثير المجتمع j.

$i = 1, 2, \ldots, r$ وأن r عدد الصفوف.

$j = 1, 2, \ldots, c$ وأن c عدد الأعمدة.

وأن $\sum_i \alpha_i = \sum_j \beta_j = 0$ ،

كما أن e_{ij} (لكل i ,j) تخضع للتوزيع الطبيعي ومستقلة عن بعضها البعض وأن معدلها يساوي صفرا ولها نفس التباين. وفي تحليل التباين الثنائي والذي افترضنا فيه عدم التفاعل بين عامل

الأعمدة وعامل الصفوف، نجد أن التغير الكلي يقسم إلى ثلاثة أقسام أو إلى ثلاثة مركبات، حيث تعزى المركبة الأولى لعامل الصفوف والثانية لعامل الأعمدة أما المركبة الثالثة فهي تعزى لعامل الصدفة. وفيما يلي تعريف بكل مركبة.

(1) المجموع الكلي للمربعات (SST): وهو مجموع مربعات انحرافات المشاهدات عن وسطها الحسابي، ويعرف بالمعادلة الآتية:

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{c}(Y_{ij} - \bar{Y})^2$$
$$= \sum_{i}\sum_{j} Y_{ij}^2 - N\bar{Y}^2$$

وله درجات حرية تساوي $rc - 1$

(2) مجموع مربعات الصفوف (SSR):

$$SSR = c\sum_{i=1}^{r}(\bar{Y}_i - \bar{Y})^2$$
$$= c\sum_{i}\bar{Y}_i^2 - N\bar{Y}^2$$

وله درجات حرية تساوي $r - 1$.

(3) مجموع مربعات الأعمدة (SSC):

$$SSC = r\sum_{j=1}^{c}(\bar{Y}_j - \bar{Y})^2$$
$$= r\sum_{j}\bar{Y}_j^2 - N\bar{Y}^2$$

بدرجات حرية تساوي $c - 1$.

(4) مجموع مربعات الخطأ (SSE):

$$SSE = \sum_{i}\sum_{j}(Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y})^2$$

وله درجات حرية تساوي $(c - 1)(r - 1)$. ومن مجاميع المربعات أعلاه يمكن كتابة ما يسمى بمتطابقة مجموع المربعات وهي:

$$SST = SSR + SSC + SSE$$

ومن هذه المتطابقة يمكن حساب SSE بطريقة أسهل بعد حساب مجاميع المربعات الأخرى حيث:

$$SSE = SST - SSR - SSC$$

وتتلخص طريقة الاختبار بالآتي:

(1) وضع الفروض: باستعمال النموذج (3) نختبر الفرضية:

$$H_0 : \alpha_i = 0, \qquad i = 1, 2, \ldots, r$$

أي أن تأثير الصفوف متساوي، مقابل الفرض:

$$H_1 : \alpha_i \neq 0$$

وهنا نرفض H_0 على مستوى معنوية α إذا كان:

$$F_1 = \frac{MSR}{MSE} \geq f_{\alpha,(r-1,(r-1)(c-1))}$$

وبنفس الطريقة نختبر الفرض العدم:

$$H_0 : \beta_j = 0, \qquad j = 1, 2, \ldots, c$$

أي أن تأثير الأعمدة واحد، مقابل الفرض:

$$H_1 : \beta_j \neq 0$$

وهنا نرفض H_0 أيضا على مستوى معنوية α إذا كان:

$$F_2 = \frac{MSC}{MSE} \geq f_{\alpha,(c-1,(r-1)(c-1))}$$

ويمكن تلخيص جميع النتائج السابقة في جدول تحليل التباين الثنائي الآتي:

جدول (رقم 31)

جدول تحليل التباين الثنائي

(بدون تفاعل)

مصدر التغير	مجموع المربعات	درجات الحرية	معدلات المربعات	قيمة F_c
بين معدلات الصفوف	SSR	$r-1$	$MSR=\frac{SSR}{r-1}$	$F_1=\frac{MSR}{MSE}$
بين معدلات الأعمدة	SSC	$c-1$	$MSC=\frac{SSC}{c-1}$	$F_2=\frac{MSC}{MSE}$
الخطأ	SSE	$(r-1)(c-1)$	$MSE=\frac{SSE}{(r-1)(c-1)}$	
المجموع الكلي	SST	$rc-1$		

- **مثال (77):** استعملت 5 أنواع من التربة على أربعة أنواع من بذور الذرة زرعت في 20 قطعة أرض تم اختيارها عشوائيا فكانت النتائج كالتالي:

نوع التربة	نوع حبوب الذرة			
	I	II	III	IV
1	12	15	10	14
2	15	19	12	11
3	14	18	15	12
4	11	16	12	16
5	16	17	11	14

والمطلوب تحت مستوى معنوية 5% التأكد من:

(1) وجود أو عدم وجود فروق بين تأثير الأنواع المختلفة من التربة.

(2) وجود أو عدم وجود فروق بين تأثير الأنواع المختلفة من بذور الذرة.

الحل: الجدول التالي يبين حساب معدلات الصفوف ومعدلات الأعمدة والمعدل العام.

جدول (رقم 32)

أنواع التربة	نوع الحبوب				مجموع الصفوف	متوسط الصفوف $\bar{Y}_j$
	I	II	III	IV		
1	12	15	10	14	51	12.75
2	15	19	12	11	57	14.25
3	14	18	15	12	59	14.75
4	11	16	12	16	55	13.75
5	16	17	11	14	58	14.50
المجموع الأعمدة	**68**	**85**	**60**	**67**	**المجموع العام: 280**	
متوسط الأعمدة $\bar{Y}_i$	13.6	17	12	13.4	المتوسط العام ($\bar{Y}$) = 14	

أما بالنسبة لمجاميع المربعات فهي:

$$(1)\quad SST = \sum_{i=1}^{5}\sum_{j=1}^{4} Y_{ij}^2 - N\bar{Y}^2$$

$$= (12)^2 + (15)^2 + \cdots + (16)^2 + (14)^2 - 20(14)^2$$

$$= 4044 - 3920 = 124$$

$$(2)\quad SSR = 4\sum_{i=1}^{5} Y_i^2 - N\bar{Y}^2$$

$$= 4\{(12.75)^2 + \ldots + (14.5)^2\} - 20(14)^2$$

$$= 3930 - 3920 = 10$$

$$(3)\quad SSC = 5\sum_{j=1}^{4} Y_j^2 - N\overline{Y}^2$$
$$= 5\{(13.6)^2 + (17)^2 + (12)^2 + (13.4)^2\} - 20(14)^2$$
$$= 3987.6 - 3920 = 67.6$$

$$(4)\quad SSE = SST - SSR - SSC$$
$$= 124 - 10 - 67.6 = 46.4$$

ومن النتائج السابقة

نستطيع أن نبني جدول تحليل التباين الثنائي التالي.

جدول (رقم 33)

جدول تحليل التباين الثنائي *Two- way ANOVA Table*

مصدر التغير	مجموع المربعات	درجات الحرية	معدلات المربعات	قيمة F_c
بين معدلات الصفوف	$SSR = 10$	$r - 1 = 4$	$MSR = \frac{10}{4} = 2.5$	$F_1 = \frac{MSR}{MSE} = 0.47$
بين معدلات الأعمدة	$SSC = 67.6$	$c - 1 = 3$	$MSC = \frac{67.6}{3} = 22.53$	$F_2 = \frac{MSC}{MSE} = 5.82$
الخطأ	$SSE = 46.4$	$(r-1)(c-1) = 12$	$MSE = \frac{46.4}{12} = 3.87$	
المجموع الكلي	$SST = 120$	$N - 1 = 19$		

وللتأكد من وجود فروق بين الأنواع المختلفة من التربة، فإننا نجري اختبار F كما يلي:

(1) **نختبر الفرض العدم الآتي:**

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$$

مقابل الفرض البديل:

$$H_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \alpha_4 \neq \alpha_5$$

(2) القيمة الجدولية لـ f هي:

$$f_{0.05,(4,12)} = 3.26$$

(3) القيمة المحسوبة لـ F_1 من جدول تحليل التباين (رقم 33) هي 0.47.

(4) لأن قيمة F_1- المحسوبة أقل من واحد وهي بالتالي أصغر من قيمة $f-$ الجدولية، نقبل الفرض العدم القائل بعدم وجود اختلافات معنوية بين الأنواع المختلفة للتربة على مستوى معنوية 5%.

وأخيرا ولكي نختبر وجود فروق بين الأنواع المختلفة من حبوب الذرة نقوم مرة أخرى بإجراء اختبار F كما يلي:

(1) نختبر الفرضية الأولى:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

مقابل الفرضية الثانية:

$$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

(2) قيمة $f-$ الجدولية هي:

$$f_{0.05,(3,12)} = 3.49$$

(3) قيمة F_2- المحسوبة من جدول (رقم 33) هي 5.82.

(4) وبما أن قيمة F_2- المحسوبة أكبر من قيمة $f-$ الجدولية، فإننا نرفض الفرض العدم ونقبل الفرض البديل القائل أن هناك فروقا معنوية بين الأنواع المختلفة من حبوب الذرة على مستوى معنوية 5% .

تمـــارين

(1) الجدول التالي يمثل نتائج ثلاثة أنواع من البذور لنبات معين

النوع الأول	99	96	95	98	97
النوع الثاني	70	65	60	65	65
النوع الثالث	66	75	87	69	89

والمطلوب باحتمال 99%:

(1) اختبار هل أن أنواع البذور تختلف معنويا؟

(2) استعمل اختبار t– المتعدد للمقارنة بين معدلات أنواع البذور الثلاثة.

(2) استعملت (6) أنواع من الأغذية في تجربة لدراسة تأثير بعض أنواع الأغذية على نمو 30 رأسا من رؤوس الماشية والتي قسمت إلى (6) مجاميع كل مجموعة مكونة من 5 رؤوس. والجدول التالي يبين الزيادة في الوزن بعد مضي مدة معينة من تناولهم للأغذية.

أنواع الأغذية	الزيادة في الوزن				
I	20	33	27	32	33
II	18	25	28	25	24
III	17	19	9	12	16
IV	21	21	21	19	19
V	14	14	12	12	14
VI	17	19	19	17	21

والمطلوب اختبار الفرضية القائلة بأن تأثير أنواع الأغذية الستة تأثير متساوي تحت مستوى احتمال 5%.

(3) أكمل جدول تحليل التباين التالي إذا علمت بأن عدد المعالجات يساوي (4) وعدد المشاهدات في كل معالجة يساوي 10

$S.O.V$	$d.f$	SS	MS	F
بسبب المعالجات			100	
بسبب الخطأ				
المجموع الكلي		500		

(4) أراد أحد الباحثين دراسة تأثير (5) مبيدات على نمو نوع معين من الفطريات، فاستخدم لهذا الغرض 50 طبقا وزعت عشوائيا إلى (5) مجاميع متساوية فحصل على النتائج الآتية:

$$MSR = 500, \quad SSE = 900$$

والمطلوب:

(1) أكمل جدول تحليل التباين.

(2) هل ترى أن هناك فروقا معنوية بين المبيدات تحت مستوى احتمال 5%.

(5) لمعرفة تأثير نوع البذور على الإنتاج أخذت 10 أنواع من البذور وزعت على 10 مجموعات من الأرض كل مجموعة تحتوي على 5 قطع، وكانت متوسطات القطع كالتالي:

رقم القطعة	1	2	3	4	5	6	7	8	9	10
المتوسط	97	65	70	59	62	82	73	80	85	90

وبإجراء تحليل التباين تحت مستوى معنوية 5% وجد أن الاختلاف ذو دلالة بين معدلات الإنتاج للأنواع المختلفة للبذور. فإذا علم من جدول تحليل التباين أن: $MSE =$ 9، المطلوب: إجراء اختبار $L.S.D$، اختبار شفييه واختبار t – المتعدد لمعرفة الأنواع المتسببة بمعنوية F.

(6) استعملت ثلاثة أنواع من الأسمدة على أربعة أنواع من الذرة فأعطت النتائج الآتية:

أنواع الذرة / أنواع السماد	*I*	*II*	*III*	*IV*
الأول	73	70	76	62
الثاني	68	68	70	59
الثالث	65	63	59	60

والمطلوب:

(1) بناء جدول تحليل التباين.

(2) اختبار فرضية أن أنواع الذرة لها نفس التأثير على الإنتاج على مستوى معنوية 5% بافتراض أن قطع الأرض متساوية الخصوبة.

(3) اختبر فرضية أن أنواع الأسمدة لها نفس التأثير على الإنتاج.

(7) كانت تكاليف ثلاثة أنواع من أنظمة التدفئة للمنازل الحديثة كما يلي:

التدفئة بالغاز	التدفئة بالكهرباء	التدفئة بالنفط
6.1	8.2	5.2
7.2	7.5	6.1
5.4	8.7	5.7
5.8	7.4	5.4
6.3	7.8	6.2
6.7	9.2	
7.4		
6.4		

(1) هل يوجد اختلاف ذو دلالة بين معدلات تكاليف الأنواع الثلاثة من أنظمة التدفئة على مستوى دلالة 5%.

(2) استعمل اختبار شفييه للمقارنة بين معدلات تكاليف أنظمة التدفئة الثلاثة.

(8) أخذت عينات من الهواء عشوائيا في ثلاثة أوقات مختلفة من اليوم وذلك لمقارنة مدى تلوث الهواء من ثلاثة مدن مختلفة من مدن العراق، فكانت النتائج كالتالي:

	ذي قار	البصرة	بغداد
9 صباحا	15	18	6
12 ظهرا	19	22	10
4 مساءا	20	23	20

والمطلوب:

(أ) إجراء تحليل التباين.

(ب) هل هناك تفاوت في درجة تلوث الهواء عبر الأوقات الثلاثة؟

(ج) هل هناك تفاوت في درجة تلوث المدن الثلاثة؟

الفصل السابع

تحليل السلاسل الزمنية

تحليل السلاسل الزمنية

Time Series Analysis

(1.7) مقدمة (*Introduction*)

إذا بحثت في ظاهرة معينة خلال فترة طويلة نسبيا من الزمن وجدت أن قيم هذه الظاهرة تتغير بمضي الزمن، وهي في تغيرها هذا تخضع لعدة مؤثرات بعضها منتظم أي ناشئ عن عوامل يمكن حصرها ومعرفة درجة تأثيرها على الظاهرة بشكل تقريبي، والآخر عبارة عن مؤثرات عرضية فجائية تؤثر على الظاهرة حيثما اتفق بالصعود أحيانا وبالهبوط أحيانا أخرى دون أن تتبعها أي قاعدة يمكن جعلها موضع دراسة علمية معينة.

وللسلسلة الزمنية استعمالات كثيرة من أهمها التنبؤ عن المستقبل باستعمال البيانات الإحصائية المتوفرة عن الماضي، واكتشاف الدورات التي تتكرر في البيانات كازدحام حركة المرور في ساعات الذروة الأمر الذي يلقي الضوء على ما ينتج من حوادث ترافق هذه الظاهرة. ولعمل كل ذلك نحتاج إلى تحليل السلسلة الزمنية المعطاة إلى مكوناتها الأساسية.

(2.7) مكونات السلسلة الزمنية

يمكن تمييز أربعة أنواع من التغيرات التي تطرأ على قيم الظواهر والتي يطلق عليها مكونات أو مركبات السلسلة الزمنية. **وهذه المكونات هي:**

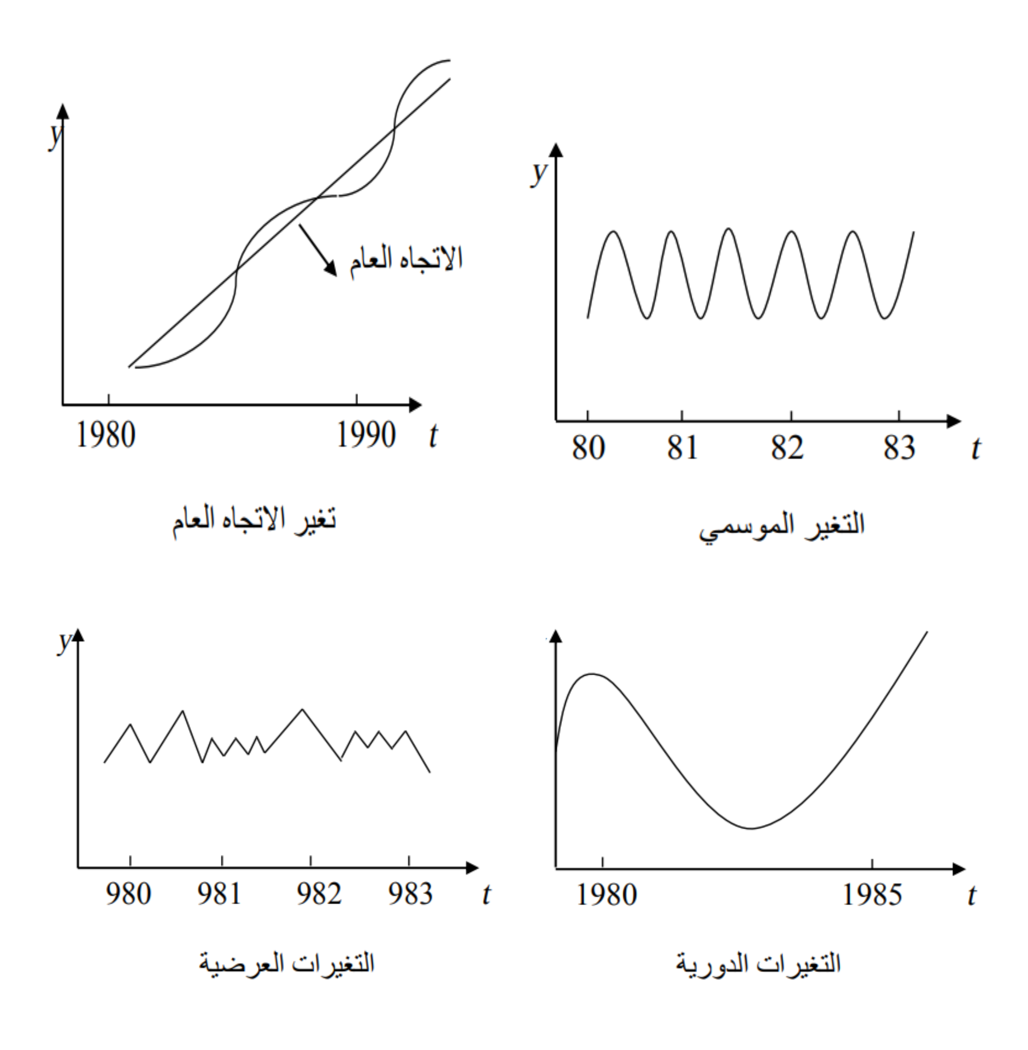
(1) التغيرات الاتجاهية *Secular Trend (T)*

(2) التغيرات الموسمية *Seasonal Variation (S)*

(3) التغيرات الدورية *Cyclical Variation (C)*

(4) التغيرات العرضية الغير منتظمة *Irregular Variations (I)*

والشكل الآتي يمثل الشكل البياني لكل واحدة من مكونات السلسلة الزمنية الأربعة.



شكل (رقم 25)

وفيما يلي شرح مختصر لطبيعة كل من هذه التغيرات.

أولا: التغيرات الاتجاهية

إذا بحثت في ظاهرة معينة كظاهرة استهلاك السكر مثلا في أي منطقة من مناطق العالم خلال فترة طويلة من الزمن فإن أول ما يلفت النظر في التغيرات التي تطرأ على قيمة هذه الظاهرة هو النمو الطبيعي المتدرج عاما بعد عام حيث تجد أن الاستهلاك الآن أضعافا مضاعفة لما كان عليه منذ خمسين أو مائة سنة. وفي ذلك أسباب طبعا، فقد يعود السبب إلى زيادة الإنتاج العالمي منه لتقدم الطرق الفنية في إنتاجه أو إلى زيادة المساحة المخصصة لزراعة قصب السكر والبنجر أو أخيرا إلى زيادة السكان وتحسين مستوى المعيشة الذي يتماشى مع درجة التقدم حيث أصبح السكر من ضروريات الحياة لكل طبقة من طبقات المجتمع بعد أن كان استهلاكه قاصرا على الطبقات الغنية فقط.

غير أن هناك ظواهر أخرى تأخذ في الانكماش التدريجي عاما بعد عام، وهي في انكماشها هذا خاضعة لما يطرأ على حياة الإنسان من تطور فيغير الطرق التي يتبعها في إشباع حاجاته. ومن هذه الظواهر، ظاهرة تناقص استهلاك الحرير الطبيعي بعد ظهور الحرير الصناعي وانتشار استعماله بين سكان العالم، أو كظاهرة انكماش استهلاك الكيروسين في كل الدول المتحضرة بعد أن حل الغاز والكهرباء محله كوقود للمنازل والمصانع مثلا.

وما يهتم به الإحصائي في دراسة الاتجاه العام للظواهر هو محاولة الوصول إلى قاعدة منظمة يمكن بواسطتها وصف سير الظاهرة في الظروف العادية وقياس مقدار الانحرافات عن هذا الاتجاه العام توطئة لمعرفة أسبابها.

ثانيا: التغيرات الموسمية

وهي التغيرات التي تحدث بانتظام في وحدات زمنية متعاقبة كشهر معين من أشهر السنة أو يوم معين من كل شهر أو حتى ساعة معينة من كل يوم. فلو بحثت مثلا في

ظاهرة استهلاك اللحوم في العراق مثلا لوجدت بلا شك أن هناك تغيرا موسميا يتخذ شكل زيادة كبيرة في الاستهلاك في عيد الأضحى المبارك ويتكرر هذا التغير بانتظام في نفس الموعد من كل سنة. كما أن حركة سحب الودائع من البنوك تزداد وتتكرر في أول كل شهر وهو الموعد الذي يحل فيه دفع مرتبات الموظفين وإيجارات المنازل والدفعات الشهرية المختلفة.

ثالثا: التغيرات الدورية

من الملاحظ أن بعض الأنشطة الاقتصادية المسيرة بنظام معين تحل بها فترات من الرخاء والكساد تتعاقب في شيء من الانتظام كل عشر سنوات تقريبا. ويطلق الاقتصاديون على هذه الفترات اسم الدورة التجارية. وما يهمنا من الوجهة الإحصائية هو تأثير هذه الدورة على قيم الظواهر التي تكون موضع بحثنا. فإذا كنا نبحث في ظاهرة استهلاك السكر في العراق ووجدنا أنه على الرغم من أن اتجاهها العام هو نحو الصعود، إلا أن الهبوط الكبير الذي طرأ عليها في سنة 1931 ما هو إلا نتيجة لتأثير الكساد التجاري الذي حل بالبلاد في تلك السنة، إذا ما عرفت فترة الكساد على أنها الفترة التي يقل خلالها الطلب الإجمالي على كافة السلع والخدمات. هذا النوع من التغير في قيم الظواهر هو المقصود بالتغيرات الدورية، مع الأخذ بنظر الاعتبار أن التغيرات الدورية لا تحدث في فترات متساوية تماما مثل التغيرات الموسمية. وإن كان حدوثها يتكرر بانتظام ولكن في فترات أطول كثيرا من فترة الدورة الموسمية.

رابعا: التغيرات العرضية

وهي تلك التغيرات التي تطرأ على قيم الظواهر أما نتيجة لتلك العوامل غير القابلة للتحديد (والتي نطلق عليها اسم عوامل المصادفة) أو لحدوث طارئ مفاجئ كالحرب أو الإضراب العام أو الفيضانات الخطيرة أو أي عامل من العوامل الطبيعية التي تؤثر على الإنتاج.

والآن إذا تأملنا في قيمة أي ظاهرة ككل لوجدنا أن هذا الكل هو عبارة عن محصلة مجموعة التغيرات أو المؤثرات السابق ذكرها والتي تشترك جميعها في تكوين هذه القيمة، وبالتالي فإن قيمة الظاهرة (Y) كما نشاهدها هي:

$$Y = T \times S \times C \times I$$

وهذا هو النموذج الإحصائي الذي يمثل السلسلة الزمنية. وعلى ضوء هذه المعلومات نستطيع أن نقول أن المقصود بتحليل السلسلة الزمنية هو رد القيمة الكلية للظاهرة إلى العناصر المكونة لها، أو بعبارة أخرى عزل كل من المؤثرات الاتجاهية والموسمية والدورية والعرضية على حدة لمعرفة مدى تأثير كل منها على قيمة الظاهرة. وفيما يلي سنتناول الطرق المختلفة لحساب هذه التغيرات.

(3.7) طرق تعيين الاتجاه العام

إن أبسط طريقة لتعيين الاتجاه العام هي أن نرصد قيم الظاهرة في الوحدات الزمنية المتتالية (أي الزمن والمشاهدات) على رسم بياني ثم نمهد يدويا الخط البياني الناتج بطريقة تقديرية.

هذا التمهيد يستبعد أثر التقلبات الدورية والموسمية والعرضية (والتي تظهر في الرسم على شكل ذبذبات وتعاريج) ويجعل الخط البياني الذي يمثل الظاهرة عبارة عن شكل هندسي منتظم (قد يكون مستقيما أو منحنى) يصف الاتجاه العام لتغير الظاهرة. وعلى الرغم من أن هذه الطريقة تمتاز بالسهولة والسرعة إلا إنها تتوقف كليا على التقدير الشخصي دون أن تستند إلى أي أساس علمي. **إما الطرق الدقيقة لإيجاد الاتجاه العام فهي:**

(1) طريقة المتوسطات المتحركة *Moving Averages Method*

(2) طريقة المربعات الصغرى *Least Square Method*

أولا: طريقة المتوسطات المتحركة

عند رسم المنحنى البياني الذي يمر بنقاط الزمن والقيمة نحصل على خط غير أملس، وذلك نتيجة للتقلبات والتغيرات المتعددة التي تطرأ على قيمة الظاهرة. ولذلك إذا أخذنا متوسط قيم الظاهرة في عدة سنين فإن التقلبات المتضادة الاتجاه تميل إلى التلاشي بتعادل الانحرافات السالبة والموجبة عن هذا المتوسط والذي يعتبر بمثابة القيمة الاتجاهية للظاهرة. وأساس طريقة المتوسطات المتحركة هو اختيار فترة زمنية يعتبر الإحصائي أن متوسط قيم الظاهرة خلالها يكون خاليا من التقلبات بأنواعها. ويتوقف اختيار هذه الفترة في الغالب على طول فترة التقلبات الدورية كما تظهر في الرسم البياني الذي يعده الباحث*. فإذا كان الخط البياني الناتج من رصد قيم الظاهرة يتكون من موجات تمتد كل منها على فترة طولها ثلاث سنوات، فإننا نختار متوسط فترته ثلاث سنوات، وإذا كانت الموجات تمتد على فترات طول كل منها خمسة سنوات فإننا نختار متوسط فترته خمس سنوات وهكذا. **فإذا انتهينا من تحديد فترة المتوسط فإننا نحسب المتوسطات المتحركة كما يلي:**

(أ) المتوسطات المتحركة التي تشمل فترتها على عدد فردي من السنين: إذا أردنا حساب متوسط متحرك فترته 3 سنوات **فإننا نقوم بالآتي:**

(1) نجمع قيم الظاهرة في الثلاث سنوات الأولى ونكتب هذا المجموع أمام السنة الوسطى أي السنة الثانية في هذه الحالة. ثم نترك السنة الأولى ونأخذ مجموع السنوات الثلاث التالية (الثانية والثالثة والرابعة) ونكتب هذا المجموع مقابل

---

* في العادة لا تزيد فترة تكرار التقلبات الموسمية عن السنة، فهي تحدث في شهر أو فصل معين من كل سنة أو في فترة أقل من ذلك. ولهذا فإن أثر التقلبات الموسمية لا يظهر إذا أخذنا القيم السنوية للظاهرة. ولأننا عند حساب الاتجاه العام نأخذ عادة القيم السنوية للظاهرة لذلك فإن التقلبات التي يتعين استبعادها هي التقلبات الدورية والعرضية. ولأن التقلبات العرضية تحدث عشوائيا دونما قاعدة، فإن متوسط أي عدد من السنين كفيل بإزالتها.

السنة الوسطى لهذه المجموعة أي مقابل السنة الثالثة. ونستمر في هذه العملية حتى يكون لدينا مجموع متحرك أمام كل سنة ما عدا السنة الأولى والأخيرة.

(2) نقسم كل مجموع متحرك على ثلاثة فنحصل على متوسطات متحركة لكل ثلاث سنوات. ثم نضع قيمة المتوسط المتحرك أمام المجموع المتحرك أمام المجموع المتحرك كما يتضح من المثال التالي.

- **مثال (78):** الجدول التالي يمثل نفقات مصنع بآلاف الدولارات من عام 67 – 77. والمطلوب حساب معدل متحرك بطول 3 سنوات.

جدول (رقم 34)

نفقات مصنع بآلاف الدولارات من عام

1977 – 1967

السنة	1967	68	69	70	71	72	73	74	75	76	77
النفقات	336	380	340	340	376	334	352	385	340	361	400

الحل: نحسب كافة المعادلات المتحركة بطول 3 وذلك بأن نجد:

$X_t + X_{t+1} + X_{t+2} /3$ لكل قيم t الممكنة. والجدول التالي يبين خطوات حساب المجاميع المتحركة والمتوسطات بطول 3.

جدول (رقم 35)

السنة	النفقات	مجموع متحرك فترته 3	متوسط متحرك فترته 3
1967	336		
68	380	1056	352
69	340	1060	353
70	340	1056	352
71	376	1050	350
72	334	1062	354
73	352	1071	357
74	385	1077	359
75	340	1086	362
76	361	1101	367
77	400		

(ب) المتوسطات المتحركة التي تشمل فترتها على عدد زوجي من السنين

إذا أردنا حساب متوسط متحرك فترته سنتان مثلا، نحسب المجاميع المتحركة لكل سنتين كالمعتاد. ونظرا لعدم وجود سنة وسطى في هذه الحالة لنضع أمامها المجموع فإننا نلجأ إلى وضع المجموع المتحرك بين السنة الأولى والثانية ثم بين السنة الثانية والثالثة وهكذا. فإذا حسبنا المتوسطات المتحركة بقسمة كل مجموع على 2 ووضعنا كل متوسط أمام المجموع المقابل له، فإن هذه المتوسطات لا تكون مقابلة للسنوات التي تكون السلسلة الزمنية وبالتالي فهي قليلة الفائدة، ذلك لأن المقصود من حساب هذه المتوسطات المتحركة هو إيجاد القيم الاتجاهية المقابلة لكل سنة. ولذلك نقوم بحساب مجموع متحرك فترته سنتان للمجموع الأول ثم قسمة المجاميع الأخيرة على 4 لنحصل على متوسطات متحركة فترتها

سنتان حيث كل متوسط منها موضوع أمام سنة معينة، كما يتضح من الجدول في المثال التالي.

- **مثال (79)**: في مثال (78)، المطلوب حساب معدل متحرك بطول 2 سنة.

الحل:

جدول (رقم 36)

إيجاد الاتجاه العام للنفقات عن طريق حساب

متوسط متحرك فترته سنتان

السنة	النفقات	مجموع متحرك فترته سنتان	مجموع متحرك فترته سنتان للمجموع السابق	متوسط متحرك فترته سنتان
1967	336			
		716		
68	380		1436	359
		720		
69	340		1400	350
		680		
70	340		1396	349
		716		
71	376		1426	357
		710		
72	334		1396	349
		686		
73	352		1423	356
		737		
74	385		1462	366
		725		

السنة	النفقات	مجموع متحرك فترته سنتان	مجموع متحرك فترته سنتان للمجموع السابق	متوسط متحرك فترته سنتان
75	340		1426	357
		701		
76	361		1462	366
		761		
77	400			

أما إذا أردنا حساب متوسط متحرك فترته (4) سنوات فإننا نتبع نفس الطريقة السابقة فنحسب أولا مجموعا متحركا فترته (4) سنوات، ثم نحسب لهذا المجموع مجموعا متحركا آخر فترته سنتان ونقسم الناتج على (8) لنحصل على المتوسط المطلوب. ومن هذا يتضح أن المجموع المتحرك الأول تكون فترته دائما مماثلة لفترة المتوسط المتحرك المراد حسابه، أما المجموع المتحرك الثاني الذي يراد به تعديل مراكز المجاميع الأولى وجعلها مقابلة للسنوات فإن فترته تكون سنتين في كل حالة. والجدول الآتي يوضح طريقة حساب متوسط متحرك معدل فترته 4 سنوات للبيانات الواردة في مثال (78).

جدول (رقم 37)

إيجاد متوسط متحرك فترته 4 سنوات

السنة	النفقات	مجموع متحرك فترته 4 سنوات	مجموع متحرك فترته سنتان للمجموع السابق	متوسط متحرك معدل فترته 4 سنوات
1967	336			
68	380			
		1396		

السنة	النفقات	مجموع متحرك فترته 4 سنوات	مجموع متحرك فترته سنتان للمجموع السابق	متوسط متحرك معدل فترته 4 سنوات
69	340		2832	354
		1436		
70	340		2826	353
		1390		
71	376		2792	349
		1402		
72	334		2849	356
		1447		
73	352		2858	357
		1411		
74	385		2849	356
		1438		
75	340		2924	366
		1486		
76	361			
77	400			

ومما يجب ملاحظته انه كلما كان عدد السنين التي ندرسها كبيرا كلما كانت القيم الاتجاهية أقرب إلى الدقة، وذلك لأن الانحرافات السالبة عن الاتجاه العام تميل في المدة الطويلة إلى التعادل مع الانحرافات الموجبة. إلى جانب ذلك أنه كلما كانت المدة طويلة كان عدد القيم الاتجاهية التي نحصل عليها كبيرا، وذلك لأن طريقة المتوسطات المتحركة لا تعطينا قيما اتجاهية لجميع السنين بل أنها تترك بعض

السنين بدون قيم اتجاهية. فلو كنا مثلا نحسب متوسطا متحركا فترته 3 سنوات لاضطررنا إلى ترك السنة الأولى والأخيرة بدون قيمة اتجاهية. **وعلى وجه العموم فإنه:**

(1) إذا كانت فترة المتوسط عددا فرديا (n مثلا)، كان عدد القيم الاتجاهية الناقصة في كل طرف من طرفي السلسلة الزمنية هو: $(n-1)/2$.

(2) إذا كانت الفترة زوجية، كان عدد القيم الناقصة في كل طرف هو: $\frac{n}{2}$.

وأخيرا وقبل أن نترك طريقة المتوسطات المتحركة لإيجاد الاتجاه العام لابد أن نجيب على السؤال الذي يقول: هل أن طريقة المتوسطات المتحركة تزيل فعلا أثر التقلبات التي تطرأ على قيمة الظاهرة أم لا؟ وهل أن حساب المتوسطات المتحركة يستبعد فعلا أثر جميع هذه التقلبات؟

إذا أردنا أن نتأكد من صحة هذا فإننا إذا أخذنا سلسلة زمنية خالية تماما من التقلبات وحسبنا لها متوسطا متحركا أيا كان فترته فإننا يجب أن نحصل على نفس القيم الأصلية للسلسلة الزمنية طالما أن ليس هناك تقلبات يزيلها حساب المتوسط. وفي الواقع إذا قمنا باختبار كهذا، فإن نتيجة الاختبار تؤيد الفرض القائل بأن حساب المتوسطات المتحركة يزيل أثر التقلبات. وللتثبت من صحة ذلك نأخذ البيانات أدناه الخالية تماما من التقلبات ونحسب لها متوسطات متحركة بطول (5) على النحو المبين في الجدول التالي.

جدول (رقم 38)

السنة	قيم الظاهرة	مجموع متحرك فترته 5	متوسط متحرك فترته 5
1970	1		
71	3		
72	5	25	5
73	7	35	7
74	9	45	9
75	11	55	11
76	13	65	13
77	15	75	15
78	17		
79	19		

ومن الجدول أعلاه يتضح أننا إذا أخذنا سلسلة زمنية خالية تماما من التقلبات وحسبنا لها متوسطا متحركا فإننا نحصل على نفس قيم السلسلة مرة أخرى. وهذه النتيجة تكون صحيحة إذا كانت قيم الظاهرة يمثلها مستقيم كما هو الحال في المثال السابق. أما إذا كان الخط البياني الذي يمثل الظاهرة منحنيا فلا شك في أن المتوسطات المتحركة تقلل من درجة انحناء المنحنى لكنها لا تصلح لإيجاد الاتجاه العام إذا كان هذا الاتجاه منحنيا.

ثانيا: طريقة المربعات الصغرى

ليس من شك في أن طريقة إيجاد الاتجاه العام بتوفيق مستقيم بطريقة المربعات الصغرى أفضل وأعم من طريقة المتوسطات المتحركة. فالطريقة الأخيرة – كما ذكرنا – لا تصلح لإيجاد الاتجاه العام إذا كان تغير الظاهرة فيه انحناء ظاهر على عكس طريقة المربعات الصغرى فهي تصلح لوصف اتجاه تغير الظواهر أيا كان شكل هذا الاتجاه. فمن المعلوم أن هناك مجموعة كبيرة من المنحنيات التي يمكن اختيار أحدها لوصف الاتجاه العام لأية ظاهرة، على أن دقة النتائج التي نحصل عليها بهذه

الطريقة تتوقف كلية على اختيار نوع المنحنى الذي يناسب تغير الظاهرة. ففي بعض الحالات يكفي مجرد النظر إلى الرسم البياني الذي يسجل قيم الظاهرة كما نشاهدها لتعيين نوع المنحنى الذي يمثلها. فقد نجد أن الخط البياني الذي يمثل تغيير الظاهرة قريب جدا من المستقيم كما هو الحال لو رصدنا على الرسم قيما تمثل ظاهرة تزداد أو تنقص بكمية ثابتة تقريبا سنة بعد أخرى.

ومن الواضح في حالة كهذه أنه يمكن إيجاد الاتجاه العام للسلسلة بتوفيق مستقيم بطريقة المربعات الصغرى. وقد نجد بمجرد النظر إلى الخط البياني الذي يمثل قيم الظاهرة أنه أقرب إلى المنحنى منه إلى المستقيم. وفي هذه الحالة نرى أن الاتجاه العام لهذه الظاهرة يصفه منحنى من الدرجة الثانية (أي قطع مكافئ) أو الثالثة أو من درجات عليا ليتم توفيقه بطريقة المربعات الصغرى كذلك.

وبالإضافة إلى المستقيم والقطع المكافئ لوصف الاتجاه العام للظاهرة. هناك مجموعة أخرى كبيرة من المنحنيات التي يمكن توفيقها للظواهر المختلفة حسب طبيعة تغير هذه الظواهر.

وسوف نتحدث في هذا البند عن توفيق المنحنيات بطريقة المربعات الصغرى عندما يكون الاتجاه خطيا مرة وغير خطي مرة أخرى.

(أ) توفيق المنحنيات الخطية

يمكن إيجاد الاتجاه العام للسلسلة الزمنية بتوفيق خط مستقيم بطريقة المربعات الصغرى. فإذا نظرنا إلى الشكل البياني الذي يمثل نفقات المصنع من السنة 1967 وحتى 1977 والواردة في مثال (76) وتبين لنا على سبيل المثال أن الاتجاه العام لهذه الظاهرة يصفه مستقيم فإن الجدول الآتي يوضح خطوات عمل توفيق مستقيم بطريقة المربعات الصغرى والذي يلخص بإيجاد خط الانحدار للبيانات المعطاة (وهو خط انحدار قيم الظاهرة (Y) على الزمن (X)).

جدول (رقم 39)

توفيق مستقيم بطريقة المربعات

الصغرى لنفقات مصنع

السنة	X	النفقات Y	X^2	XY	القيم الاتجاهية $\hat{Y}$
1967	-5	336	25	-1680	344.45
68	-4	380	16	-1520	347.27
69	-3	340	9	-1020	350.09
70	-2	340	4	-680	352.91
71	-1	376	1	-376	355.73
72	0	334	0	0	358.55
73	1	352	1	352	361.37
74	2	385	4	770	364.19
75	3	340	9	1020	367.01
76	4	361	16	1444	369.83
77	5	400	25	2000	372.65
		3944	110	310	

فمن المعروف أن معادلة المستقيم هي: $Y = \alpha + \beta X$ **، حيث:**

X = المتغير المستقل (وهو الزمن).

Y= المتغير التابع (قيمة الظاهرة).

α = قيمة Y عندما تكون $X = 0$.

β = ميل المستقيم بالنسبة للمحور السيني.

ويمكن إيجاد قيم α, β **من المعادلات الطبيعية الآتية:**

$$\sum Y = n\hat{\alpha} + \hat{\beta}\sum X_i$$
$$\sum XY = \hat{\alpha}\sum X + \hat{\beta}\sum X^2$$

والآن وبالنظر إلى جدول (رقم 39) نلاحظ أننا اختصرنا خطوات العمل كثيرا بجعل نقطة الأصل عند السنة الوسطى (1972) فأصبح $\sum X = 0$ ، وبالتالي فإن قيمة $\hat{\alpha}$ هي:

$$\hat{\alpha} = \frac{\sum Y}{n} = \frac{3944}{11} = 358.55$$

وأن قيمة $\hat{\beta}$ هي:

$$\hat{\beta} = \frac{\sum XY}{\sum X^2} = \frac{310}{110} = 2.82$$

ومنها فإن معادلة خط الاتجاه العام هي:

$$\hat{Y} = 358.55 + 2.82\,X$$

حيث نقطة الأصل عند سنة 1972. وبالتعويض عن قيم X في هذه المعادلة نحصل على القيم الاتجاهية الموضحة بالعمود الأخير في الجدول (رقم 39). فمثلا للحصول على القيمة الاتجاهية للظاهرة في سنة 1974 نجري العمل كالآتي:

$$\hat{Y}_{1974} = 358.55 + 2.82(2) = 364.19$$

❖ **حالة كون عدد السنين زوجيا**

عندما تشتمل السلسلة الزمنية على عدد زوجي من السنين فلا يمكن اختصار خطوات العمل بالطريقة السابقة، وذلك لعدم وجود سنة وسطى نعتبرها نقطة الأصل. وللتغلب على هذه الصعوبة نعتبر نقطة الأصل في منتصف المسافة الواقعة بين السنتين المتوسطتين ونجعل وحدات (X) تمثل نصف سنة كما هو مبين بالجدول الآتي الذي يوضح خطوات عمل توفيق مستقيم بطريقة المربعات الصغرى في حالة كون عدد السنين زوجيا.

جدول (رقم 40)

السنة	القيمة المشاهدة Y	X	X^2	XY	القيم الاتجاهية $\hat{Y}$
1980	137	-9	81	-1233	143.3
81	165	-7	49	-1155	151.46
82	140	-5	25	-700	159.56
83	181	-3	9	-543	167.66
84	188	-1	1	-188	175.75
نقطة الأصل	–	–	–	–	179.80
85	169	1	1	169	183.85
86	190	3	9	570	191.94
87	210	5	25	1050	200.04
88	198	7	49	1386	208.14
89	220	9	81	1980	216.23
المجموع	**1798**		**330**	**1336**	

ومن هذه الحسابات نجد أن قيمة $\hat{\alpha}$ هي:

$$\hat{\alpha} = \frac{\sum Y}{n} = \frac{1798}{10} = 179.8$$

أما قيمة $\hat{\beta}$ فهي:

$$\hat{\beta} = \frac{\sum XY}{\sum X^2} = \frac{1336}{330} = 4.048$$

وبذلك تكون معادلة خط الانحدار هي:

$$\hat{Y} = 179.8 + (4.048)X$$

حيث نقطة الأصل واقعة بين سنتي 1985 ،1984 وأن وحدات X تمثل نصف سنة.

ويمكن أن نستخدم معادلة المستقيم أو المنحنى الذي يصف الاتجاه العام للظاهرة في التنبؤ بما ستكون عليه قيمة الظاهرة في المستقبل على وجه التقريب وعلى فرض ثبات

الظروف المحيطة بهذه الظاهرة على أن لا يكون الزمن بعيدا لأكثر من نصف طول الفترة المعطاة.

وبالإضافة إلى الطريقة المختصرة لإيجاد خط الانحدار فهناك طرق أخرى لإيجاد هذا الخط كأن نقوم بإعطاء السنوات أرقاما (نسميها X) تصاعدية ابتداء بالواحد حتى آخر سنة. وبمقارنة القيم الاتجاهية التي نحصل عليها بموجب كل طريقة نجد أنها نفس القيم مع اختلاف معادلة المستقيم التي نحصل عليها بموجب كل طريقة وذلك عائد إلى اختلاف نقطة الأصل واختلاف وحدات (X) في كل طريقة من الطرق.

(ب) توفيق المنحنيات الغير خطية

ذكرنا أن هناك مجموعة كبيرة من المنحنيات ، إلى جانب المستقيم ، يمكن توفيقها للظواهر نذكر منها:

(1) توفيق منحنى من الدرجة الثانية

إذا أظهر لنا شكل الانتشار أن العلاقة بين الزمن وقيمة الظاهرة تبدو وكأنها على شكل منحنى فان طريقة المربعات الصغرى تعطي لنا المعادلات اللازمة لتقدير أحسن ثوابت للمعادلة التي تأخذ الشكل الآتي:

$$Y = a + bX + cX^2$$

حيث a, b, c ثوابت المعادلة المراد إيجاد تقدير لها من بيانات السلسلة أو الظاهرة محل الدراسة. ويمكن إيجاد قيم الثوابت أعلاه من المعدلات الطبيعية الآتية:

$$\sum Y = na + b\sum X + c\sum X^2$$
$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$
$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

وبحل هذه المعادلات يمكن الحصول على معادلة انحدار Y على X من الدرجة الثانية.

- **مثال (80):** الجدول التالي يمثل مشتريات أحد المخازن (بآلاف الدولارات) من سلعة معينة خلال السنوات من 1985 حتى 1992. والمطلوب توفيق منحنى من الدرجة الثانية لهذه الظاهرة.

السنة	1985	86	87	88	89	90	91
المشتريات	7	14	17	24	33	39	41

الحل:

الجدول التالي يوضح طريقة توفيق المنحنى الذي يصف الاتجاه العام لمشتريات المخزن من تلك السلعة.

جدول (رقم 41)

السنة	X	Y	XY	X^2	X^2Y	X^4	$\hat{Y}$
1985	-3	7	-21	9	63	81	6.65
86	-2	14	-28	4	56	16	13
87	-1	17	-17	1	17	1	19.21
88	0	24	0	0	0	0	25.28
89	1	33	33	1	33	1	31.21
90	2	39	78	4	156	16	37
91	3	41	123	9	369	81	42.65
المجموع		**175**	**168**	**28**	**694**	**196**	

ولأننا جعلنا نقطة الأصل عند السنة الوسطى 1988، لذلك فقد اختصرنا خطوات العمل كثيرا حيث أصبح كل من: $\sum X = \sum X^3 = 0$ ، وبالتعويض في المعادلات الطبيعية الثلاثة نجد أن:

$$175 = 7a + 28c$$

$$168 = 28b$$

$$694 = 28a + 196c$$

وبحل المعادلات الثلاثة نجد أن : $a = 25.28,\quad b = 6,\quad c = -0.07$ **وبالتالي فان**

معادلة المنحنى هي: $\hat{Y} = 25.28 + 6X - 0.07X^2$

حيث نقطة الأصل عند سنة 1988. وبالتعويض عن قيم X في هذه المعادلة نحصل على القيم الاتجاهية الموضحة بالعمود الأخير من جدول (رقم 41). فلكي نحصل على القيمة الاتجاهية للظاهرة في سنة 1990 نقوم بعمل الآتي:

$$\hat{Y}_{1990} = 25.28 + 6(2) - 0.07(2)^2$$
$$= 37$$

وهي القيمة الاتجاهية المبينة بالجدول مقابل سنة 1990.

(2) المنحنى اللوجستي (*Logistic Curve*)

ويعرف المنحنى اللوجستي بالمعادلة الآتية:

$$y = \frac{1}{c + ab^x}$$

حيث c, a, b هي الثوابت التي تحدد شكل المنحنى. وهذا المنحنى هو أنسب المنحنيات لوصف نمو الظواهر وعلى وجه الخصوص نمو السكان سنة بعد سنة. وكما نرى فانه ليس من السهل إيجاد خط الانحدار اللوجستي، وذلك لعدم وجود تحويلة (*Transformation*) سهلة تجعل من المعادلة اللوجستية معادلة خطية ليتم توفيقها بطريقة المربعات الصغرى. ومن الطرق المتبعة لتوفيق مثل هذا المنحنى هو بأن نأخذ ثلاثة قيم من قيم X ، y لتكوين ثلاث معادلات بثلاثة مجاهيل بعد أن نضع المتغير التابع على هيئة $1/y$ بدلا من y. وبحل هذه المعادلات آنيا وإيجاد المعالم c, b, a نحصل على خط الاتجاه العام اللوجستي.

- **مثال (81):** أوجد خط الاتجاه العام اللوجستي لمبيعات سلعة معينة للسنوات المبينة في الجدول أدناه.

جدول (رقم 42)

السنة	1979	1980	1981	1982	1983
X	-2	-1	0	1	2
المبيعات	3	22	39	46	51

الحل:

افرض إننا أخذنا قيم X الآتية: $X = -2, X = 0, X = 2$، وبتعويض هذه القيم في معادلة المنحنى اللوجستي نحصل على ثلاثة معادلات بثلاثة مجاهيل بعد أن نأخذ y على هيئة $1/y$، أي:

$$1/y = c + ab^{x}$$
$$1/3 = c + ab^{-2}$$
$$1/39 = c + ab^{0}$$
$$1/51 = c + ab^{2}$$

وبحل هذه المعادلات آنيا نجد أن: $c = 0.019$، $b = 0.145$، $a = 0.007$، **وبذلك تكون معادلة خط الانحدار اللوجستي هي:**

$$y = \frac{1}{0.019 + (0.007)(0.145)^{x}}$$

ولكي نجد القيمة التنبؤية لمبيعات السلعة في سنة 1984 ($X = 3$) بالاعتماد على هذه البيانات نجد أن:

$$\hat{y}_{1984} = \frac{1}{0.019 + (0.007)(0.145)^{3}} = 52.57$$

(3) المنحنى الذي معادلته $y = aX^b$

ويناسب هذا المنحنى الظواهر التي إذا رصدت على قياس لوغاريتمي مزدوج تظهر على شكل خط مستقيم أو قريب من المستقيم مثل ظاهرة توزيع الدخول. فإذا أخذنا لوغاريتمات الطرفين للمعادلة أعلاه تصبح معادلة الطرفين كالتالي:

$$\log y = \log a + b \log X$$

وبهذه التحويلة أصبح هذا المنحنى خطا مستقيما معادلته $y = \alpha + \beta X$ **إذا قمنا برصد لوغاريتمات القيم بدلا من القيم ذاتها على الرسم البياني.**

(4) المنحنى الأسي (*Exponential Curve*)

تستعمل السلاسل الزمنية عادة لوصف الظواهر التي تزداد أو تنقص بنسب ثابتة على مدى الزمن كظاهرة مبيعات منتج جديد أو ظاهرة النمو السكاني والمنحنى المناسب لظواهر كهذه هو المنحنى الأسي الذي يعتمد على قيم الثوابت a, b **والذي تمثله المعادلة التالية:** $y = ab^X$

وهنا إذا وقعت قيمة b **بين الصفر والواحد فان قيمة الظاهرة** y **تنقص كلما ازدادت قيمة** X**، بينما تزداد قيمة** y **بزيادة** X **عندما تكون** b **أكبر من الواحد. إما** a **فهي المسافة بين نقطة الأصل ونقطة تقاطع المستقيم مع المحور** y**. ومرة أخرى إذا أخذنا لوغاريتمات الطرفين فان معادلة المنحنى ألأسي تصبح على الشكل التالي:**

$$\log y = \log a + X \log b$$

ليكون المنحنى خطا مستقيما توفقه طريقة المربعات الصغرى.

- **مثال (82): الجدول التالي يمثل مبيعات سلعة معينة اعتبارا من السنة** 1979 **وحتى** 1983**. والمطلوب إيجاد خط الاتجاه العام للظاهرة بطريقة المربعات الصغرى.**

جدول (رقم 43)

السنة	1979	1980	1981	1982	1983
المبيعات	1	3	6	14	41

الحل:

إذا قمنا بتمثيل هذه البيانات بيانيا لرأينا أن المنحنى الموفق لها هو المنحنى ألأسي.

وبالتالي فان الجدول الآتي يبين خطوات الحل.

جدول (رقم 44)

السنة	X	المبيعات y	$log\ y$	$X\ log\ y$	X^2	$\hat{y}$
1979	-2	1	0	0	4	1.06
80	-1	3	0.477	-0.477	1	2.59
81	0	6	0.778	0	0	6.35
82	1	14	1.146	1.146	1	15.56
83	2	41	1.613	3.226	4	38.12
المجموع			4.014	3.895	10	

ولأن $\sum X = 0$ فان:

$$\log a = \frac{\sum \log y}{n} = 4.014/5 = 0.8028$$

$$\log b = \frac{\sum X \log y}{\sum X^2} = 3.895/10 = 0.3895$$

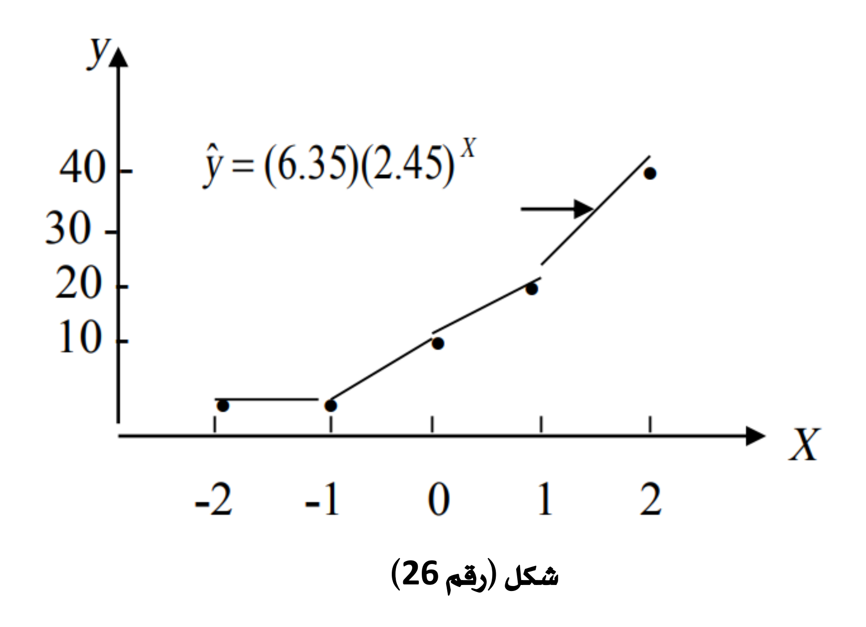
ولإيجاد تقديرات المربعات الصغرى لكل من a, b فإننا نوجد اللوغاريتم العكسي (*Antilog.*) لهذه القيم لكي نجد أن: $a = 6.35, \quad b = 2.45$ ، وبذلك تكون معادلة خط الاتجاه العام ألأسي هي:

$$\hat{y} = (6.35)(2.45)^X$$

ولكي نحسب القيم الاتجاهية لمبيعات سنة 1984، نعوض عن قيمة X بـ 3 فيكون:

$$\hat{y}_{1984} = (6.35)(2.45)^3 = 93.34$$

والشكل الآتي يوضح نتيجة توفيق المنحنى الأسي للمبيعات من تلك السلعة والذي يظهر أن هناك توافقا كبيرا بين القيم المشاهدة والقيم المحسوبة من معادلة المنحنى مما يدل على أن هذا المنحنى يصف الاتجاه العام للمبيعات بدقة.



شكل (رقم 26)

(4.7) طرق حساب التغيرات الموسمية

هناك طرق عدة لحساب التغيرات الموسمية غير أننا سنكتفي بشرح طريقتين منها فقط.

<u>**الطريقة الأولى**</u>: وهي تعطينا التغيرات الموسمية على شكل نسب مئوية (تسمى النسب الموسمية) تظهر بمجرد النظر إليها درجة تأثير قيمة الظاهرة بالموسم.

الطريقة الثانية: وهي تعطينا التغيرات الموسمية مقدرة بوحدات القياس العادية. وفيما يلي شرح لهاتين الطريقتين.

الطريقة الأولى: إيجاد النسب الموسمية بطريقة المتوسطات البسيطة

افرض أننا نريد معرفة اليوم الذي تزداد فيه حركة النقل عادة على الحافلات عن المتوسط في مدينة معينة أو اليوم الذي تقل فيه هذه الحركة عن المتوسط في تلك المدينة. فإذا اعتبرنا أن جملة الإيرادات المتحصلة يوميا من بيع التذاكر تعبر بدقة عن حركة النقل بالحافلات، فان أول طريقة تتبادر إلى الذهن لمقارنة حركة النقل في الأيام المختلفة هي أن نحسب متوسط الإيرادات المتحصلة في كل يوم من أيام الأسبوع على حدة خلال فترة طويلة نسبيا من الزمن ثم نقارن هذه المتوسطات اليومية. ولكي تكون المقارنة سهلة وواضحة يجب أن نتخذ أساسا نعتبره يمثل حركة النقل العادية. فإذا كان المتوسط اليومي أكبر من هذا الأساس ظهر لنا في الحال أن تأثير الموسم صعوديا في هذا اليوم. وبالمثل إذا كان المتوسط اليومي أقل من هذا الأساس دل ذلك على أن قيمة الظاهرة تقل بتأثير الموسم في هذا اليوم.

ولقد وجد أن أفضل طريقة لاختيار هذا الأساس هي حساب الوسط الحسابي للمتوسطات اليومية واعتباره ممثلا لحركة النقل العادية. وفي الحقيقة فان طريقة المتوسطات البسيطة لإيجاد النسب الموسمية ما هي إلا حساب متوسط كل يوم كنسبة مئوية من الوسط الحسابي للمتوسطات اليومية الذي اعتبرناه أساسا للمقارنة. **ومما سبق يتضح أن خطوات العمل اللازمة لإيجاد النسب الموسمية بهذه الطريقة هي:**

(1) حساب المتوسطات اليومية إذا كان التأثير الموسمي يظهر في يوم معين من كل أسبوع، أو المتوسطات الشهرية إذا كان التأثير الموسمي يظهر في شهر معين من كل سنة أو أي وحدة زمنية أخرى يظهر فيها التأثير الموسمي.

(2) حساب الوسط الحسابي للمتوسطات اليومية (أو الشهرية ... الخ).

(3) حساب كل متوسط من المتوسطات اليومية (أو الشهرية) كنسبة مئوية من الوسط الحسابي لهذه المتوسطات لتكون هذه النسب المئوية هي النسب الموسمية المطلوبة.

- **مثال (83):** المطلوب حساب النسب الموسمية اليومية لإيرادات التذاكر في خط حافلة معين في مدينة معينة.

جدول (رقم 45)

إيجاد النسب الموسمية لإيرادات

تذاكر خط الحافلة

رقم الأسبوع	الإيرادات اليومية لأقرب دولار واحد							
	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة	
1	1052	942	762	678	767	786	931	
2	831	1137	944	739	709	721	694	
3	760	935	809	725	730	747	721	
4	838	843	790	746	742	781	727	
5	876	958	795	735	733	744	746	
6	825	895	755	780	742	787	704	
7	880	996	794	784	750	761	720	
المجموع	6062	6706	5649	5187	5173	5327	5243	
المتوسطات اليومية	866	958	807	741	739	761	749	5621
النسب الموسمية اليومية	107.85	119.3	100.5	92.28	92.03	94.77	93.28	700

مع ملاحظة أن:

1) متوسط المتوسطات اليومية = $\frac{5621}{7} = 803$

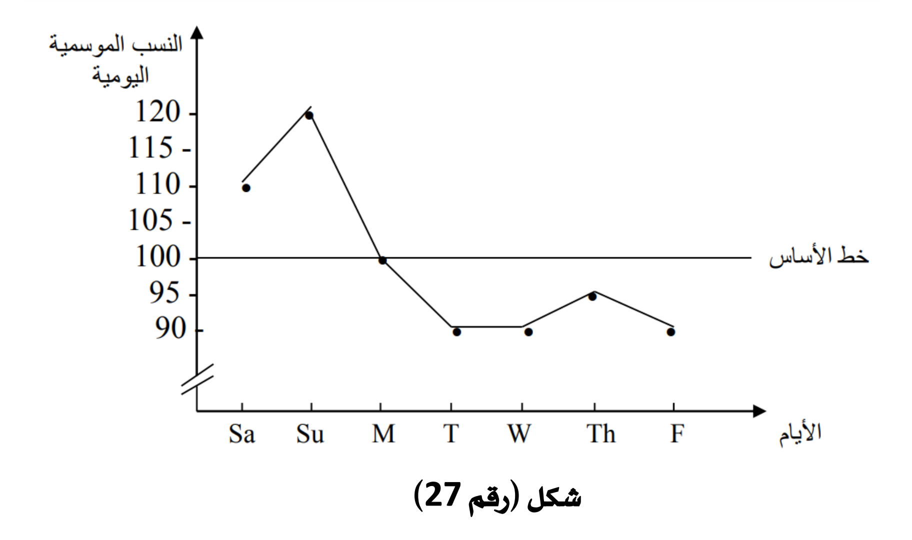
2) النسبة الموسمية لأي يوم من أيام الأسبوع تساوي المتوسط اليومي مقسوما على متوسط المتوسطات مضروبا في 100. أي أن النسبة الموسمية ليوم السبت مثلا هي:

$$\frac{866}{803} \times 100 = 107.85$$

ومن الجدول السابق يتضح أن يوم الأحد هو اليوم الذي تبلغ فيه إيرادات التذاكر في خط الحافلة حدها الأقصى. إما يوم الاثنين فان حركة النقل فيه عادية لأن الإيرادات التي تحصل عادة في هذا اليوم قريبة جدا من المتوسط (وهو 100). أما بقية أيام الأسبوع (فيما عدا السبت) فان الحركة فيها أقل من المتوسط. والشكل التالي يوضح هذه النسب الموسمية اليومية بيانيا.

النسب الموسمية اليومية لإيرادات

التذاكر في خط الحافلة



شكل (رقم 27)

❖ استبعاد الاتجاه العام من النسب الموسمية

إذا كانت قيم الظاهرة خاضعة لتأثير الاتجاه العام كما هو المعتاد فلا بد من استبعاد هذا الاتجاه عند حساب النسب الموسمية. ذلك لأن بعض هذه النسب تكون أكثر مما يجب أن تكون إذا كان الاتجاه العام للظاهرة هو نحو الصعود. أما إذا كان اتجاه الظاهرة نحو الهبوط فان بعض هذه النسب تكون أقل مما يجب أن تكون. فإذا لم يستبعد أثر الاتجاه العام فقد يبدو أن هناك تقلبات موسمية حتى إذا لم يكن هناك مؤثر موسمي حقيقي. ويمكن استبعاد تأثير الاتجاه العام من النسب الموسمية بطرق عدة وسوف نقوم بشرح واحدة منها وهي طريقة استبعاد الاتجاه قبل حساب المتوسطات.

❖ استبعاد الاتجاه قبل حساب المتوسطات

تتلخص هذه الطريقة في استبعاد الاتجاه العام من القيم الفعلية للظاهرة Y قبل اتخاذ أية خطوة لحساب النسب الموسمية. فإذا كنا نبحث في القيم اليومية لظاهرة معينة خلال سبعة أسابيع (وهنا يكون عدد القيم الفعلية للظاهرة 49، كما هو الحال في الأرقام الخاصة بإيرادات تذاكر خط الحالفة المبينة في جدول (رقم 45)) ، فان علينا أن نعين 49 قيمة اتجاهية بإحدى الطرق التي ذكرناها قبلا لإيجاد الاتجاه العام. ثم نحسب كل قيمة فعلية كنسبة مئوية من القيمة الاتجاهية المناظرة لها، لنكون بذلك قد استبعدنا أثر الاتجاه العام من القيم الفعلية للظاهرة أي أن:

القيمة الفعلية مجردة (مخلصة) من أثر الاتجاه العام = القيمة الفعلية للظاهرة مقسوما على القيمة الاتجاهية مضروبا في 100.

بعد ذلك نحسب المتوسطات الشهرية للقيم الفعلية المخلصة من أثر الاتجاه العام ومن ثم نستخرج النسب الموسمية كالمعتاد.

<u>الطريقة الثانية</u>: إيجاد التقلبات الموسمية بدلالة الوحدات المطلقة

إن الطريقة الأولى السابقة هي أفضل طريقة لمعرفة مدى تأثر الظاهرة بالموسم إلا أن مشقة العمل الحسابي اللازم لاستخراجها يجعلنا نكتفي أحيانا بالتعبير عن التقلبات الموسمية بدلالة الوحدات المطلقة. إلى جانب ذلك، فان هذه الطريقة مهمة إذا ما أردنا معرفة القيم الفعلية للتقلبات الموسمية. وإذا ما قمنا بعزل التقلبات الموسمية وذلك بحساب انحرافات القيم المشاهدة للظاهرة عن الاتجاه العام (على فرض عدم وجود تأثير دوري)، فان هذه الانحرافات عبارة عن مزيج من التقلبات الموسمية والعرضية. ولكي نستبعد التقلبات العرضية نأخذ المتوسط اليومي (أو الشهري أو الفصلي وذلك حسب فترة تكرار التقلبات الموسمية) للانحرافات عن الاتجاه العام، لأن أثر هذه التقلبات يتلاشى تقريبا عند حساب المتوسطات (أيا كانت فترتها) للانحرافات عن الاتجاه العام. وبناء على ذلك يمكن اعتبار هذه المتوسطات اليومية أو الشهرية تقلبات موسمية بحتة. **وفيما يلي خطوات العمل اللازمة لتعيين التقلبات الموسمية بدلالة الوحدات المطلقة:**

1) حساب القيم الاتجاهية للظاهرة بإيجاد متوسط متحرك تتوقف فترته على طول الدورة الموسمية.
2) حساب انحرافات القيم المشاهدة للظاهرة عن القيم الاتجاهية.
3) ترتيب هذه الانحرافات في شكل جدول يسمح باستخراج المتوسطات اليومية (أو الشهرية أو أي فترة زمنية أخرى) وذلك بوضع انحرافات كل يوم أو كل شهر في عمود واحد.
4) نحسب المتوسطات الشهرية للانحرافات فتكون هي القيم الفعلية للتقلبات الموسمية.

ولتوضيح هذه الخطوات عمليا نحسب التقلبات الموسمية لإيرادات تذاكر الحافلة في مثال 83.

- **مثال (84)**: لبيانات الأربع أسابيع الأولى في مثال (83) احسب:

(أ) التقلبات الموسمية في إيرادات التذاكر.

(ب) تحليل السلسلة إلى العناصر المكونة لها

الحل:

(أ) لإيجاد التقلبات الموسمية في إيرادات التذاكر نقوم بعمل الجدول الآتي:

جدول (رقم 46)

إيرادات تذاكر الحافلة اليومية خلال 4 أسابيع

الأسبوع		الإيرادات	مجموع متحرك بطول 7	متوسط متحرك بطول 7	الانحراف عن الاتجاه العام
الأول	السبت	1052			
	الأحد	942			
	الاثنين	762			
	الثلاثاء	678	5918	845.43	-167.43
	الأربعاء	767	5697	813.86	- 46.86
	الخميس	786	5892	841.71	- 55.71
	الجمعة	931	6074	867.71	63.29
الثاني	السبت	831	6135	876.43	- 45.43
	الأحد	1137	6077	868.14	268.86
	الاثنين	944	6112	858.86	85.14
	الثلاثاء	739	5775	825.00	- 86.00
	الأربعاء	709	5704	814.86	- 105.86
	الخميس	721	5502	786.00	- 65.00
	الجمعة	694	5367	766.71	- 72.71
	السبت	760	5353	764.71	- 4.71
	الأحد	935	5374	767.71	167.29

الأسبوع		الإيرادات	مجموع متحرك بطول 7	متوسط متحرك بطول 7	الانحراف عن الاتجاه العام
الثالث	الاثنين	809	5400	771.43	37.57
	الثلاثاء	725	5427	775.29	- 50.29
	الأربعاء	730	5505	786.43	- 56.43
	الخميس	747	5413	773.29	- 26.29
	الجمعة	721	5394	770.57	- 49.57
الرابع	السبت	838	5415	773.57	64.43
	الأحد	843	5427	775.29	67.71
	الاثنين	790	5461	780.14	9.86
	الثلاثاء	746	5467	781.00	- 35.00
	الأربعاء	742			
	الخميس	781			
	الجمعة	727			

لاحظ أننا اخترنا متوسطا متحركا بطول 7 لتمثيل الاتجاه العام للظاهرة. وبطرح القيم الاتجاهية من القيم الفعلية نحصل على الانحرافات عن الاتجاه العام المبينة بالعمود الأخير من الجدول. وكما ذكرنا من قبل فإن هذه الانحرافات هي عبارة عن مزيج من التقلبات الموسمية والعرضية ولكي نعزل التقلبات الموسمية على حدة نأخذ المتوسطات اليومية لهذه الانحرافات. وحتى نسهل حساب هذه المتوسطات نرتب الانحرافات عن الاتجاه العام في شكل جدولي كما هو آت.

جدول (رقم 47)

إيجاد التقلبات الموسمية لإيرادات تذاكر الحافلة

الأسبوع	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة
1				-167.43	- 46.86	-55.71	63.29
2	- 45.43	268.86	85.14	-86	-105.86	-65	-72.71
3	- 4.71	167.29	37.57	-50.29	-56.43	-26.29	- 49.57
4	64.43	67.71	9.89	-35			
المجموع	14.29	503.86	132.57	-338.72	-209.15	-147	-58.99
المتوسط	4.76	167.95	44.19	- 84.68	- 69.72	- 49	-19.66

وهذه المتوسطات المبينة في السطر الأخير من الجدول تمثل التقلبات الموسمية. ويلاحظ أنه كلما كان عدد السنوات التي تمتد خلالها السلسلة الزمنية كبيرا كلما كانت النتائج التي نحصل عليها أقرب إلى الدقة، ذلك لأننا افترضنا أن حساب متوسطات الانحرافات يستبعد أثر التقلبات العرضية ولا يكون هذا الفرض صحيحا ما لم تكن المدة طويلة حيث تميل التقلبات العرضية التي تسبب زيادة قيمة الظاهرة إلى التعادل مع تلك التي تحدث نقصا فيها.

(ب) الآن وقد حسبنا كلا من الاتجاه العام والتقلبات الموسمية يمكننا أن نحلل القيم الفعلية للظاهرة إلى العناصر التي تكونها، وتتضح نتيجة هذا التحليل في الأسبوعين الأول والثاني من الجدول الآتي.

جدول (رقم 48)

تحليل السلسلة الزمنية إلى العناصر المكونة لها

الأسبوع		القيم الفعلية	الاتجاه العام	التقلبات الموسمية	التقلبات العرضية
الأول	الثلاثاء	678	845.43	4.76	-172.19
	الأربعاء	767	813.86	167.95	-214.81
	الخميس	786	841.71	44.19	-89.90
	الجمعة	931	867.71	-84.68	147.97
الثاني	السبت	831	876.43	4.67	-50.10
	الأحد	1137	868.14	167.95	100.91
	الاثنين	944	858.86	44.19	40.46
	الثلاثاء	739	825.00	-84.68	-1.32
	الأربعاء	709	814.86	-69.72	-36.14
	الخميس	721	786.00	-49.00	-16.00
	الجمعة	694	766.71	-19.66	-53.05

بالنظر إلى الجدول السابق يلاحظ أنه لكي نحصل على قيم التقلبات العرضية ليوم الثلاثاء في الأسبوع الأول نقوم بالآتي:

التقلبات العرضية = 678 – (845.43 + 4.76)

= 172.19 -

وما قمنا به هو أننا طرحنا مجموع القيمة الاتجاهية والتقلبات الموسمية من القيمة الفعلية المناظرة لها. وهكذا بالنسبة لجميع قيم التقلبات العرضية الباقية.

❖ **استبعاد تأثير الموسم من القيم المشاهدة للظاهرة**

إن الغرض الرئيسي من حساب النسب الموسمية هو دراسة تأثير الموسم على قيم الظاهرة. هذه الدراسة ذات أهمية كبرى بالنسبة للمنتجين بوجه عام وأولئك الذين يتغير الطلب على منتجاتهم كثيرا من وقت لآخر. فحساب النسب الموسمية

يجعل من السهل على هؤلاء المنتجين إعداد العدة لمواجهة الزيادة (أو النقص) في الطلب على منتجاتهم. وما يهم المنتج هو الوقوف على الحالة الحقيقية للطلب على منتجاته. فقد تزداد إيراداته كثيرا في وقت معين من أوقات السنة ومع ذلك تكون هذه الزيادة مؤقتة تزول بزوال تأثير الموسم على الإيرادات. ولكي يقف المنتج على الحالة الحقيقية لإيراداته (والتي ستكون أساسا لقراراته وتصرفاته المتعلقة بمستقبل المشروع) فلابد أن يستبعد تأثير الموسم من هذه الإيرادات. ويتم استبعاد تأثير الموسم بقسمة القيم الفعلية للظاهرة على النسب الموسمية. فلو فرضنا مثلا أن إيرادات محل تجاري في شهر شباط هي 5000 دولار، وكانت النسب الموسمية للإيرادات في هذا الشهر هي 125% فإن الإيرادات بعد استبعاد الموسم تساوي:

$$\begin{aligned} &= 5000 \times \frac{100}{125} \\ &= 4000 \end{aligned}$$

أما إذا كانت النسبة الموسمية في هذا الشهر هي 50% فإن الإيرادات بعد استبعاد الموسم تساوي:

$$\begin{aligned} &= 5000 \times \frac{100}{50} \\ &= 10000 \end{aligned}$$

ومن جانب آخر فإنه يمكن استبعاد التقلبات الموسمية إذا كانت قيمها مقدرة بدلالة الوحدات المطلقة بطرحها مباشرة من القيم الفعلية للظاهرة.

(5.7) التغيرات الدورية

يمكن قياس تأثير التقلبات الدورية بالطريقة المعتادة وذلك بأن نستبعد الاتجاه العام والتأثيرات الموسمية على التوالي من قيم الظواهر الفعلية. ويحصل ذلك بقسمة القيمة المشاهدة للظاهرة على القيمة الاتجاهية ثم على النسبة الموسمية المناظرة لها. فإذا فرضنا أن I, C, S, T تمثل على التوالي القيم الاتجاهية والتقلبات الموسمية والدورية والعرضية فإن القيمة الفعلية Y تساوي:

$$Y = T \times S \times C \times I$$

وبالقسمة على T نجد أن:

$$\frac{Y}{T} = S \times C \times I$$

ثم بالقسمة مرة أخرى على S نجد أن:

$$\frac{Y}{T \times S} = C \times I$$

ولا تبقى سوى التقلبات العرضية والتي بطبيعتها لا يمكن حصرها أو استبعادها فإذا فرضنا أنها تحدث في اتجاهات مضادة كما هو الغالب فإنه يمكن استبعادها بأخذ متوسطات متحركة للناتج الأخير بعد استبعاد الاتجاه العام وتأثير الموسم.

- **مثال (85):** إذا فرضنا أن هناك تغيرات دورية في بيانات المثال (83)، فإن المطلوب حساب النسب الدورية للأسبوع الثاني فقط. (انظر جدول (رقم 45 & 46)).

جدول (رقم 49)

إيجاد النسب الدورية باستبعاد الاتجاه العام والتقلبات الموسمية

الأسبوع		القيم الفعلية $Y=T.S.C.I$	القيم الاتجاهية T	النسب الموسمية S	النسب الدورية $(\frac{Y}{T.S})\times 100$
الثاني	السبت	831	876.43	1.0785	86.92
	الأحد	1137	868.14	1.1930	109.78
	الاثنين	944	858.86	1.005	109.37
	الثلاثاء	739	825.00	0.9228	97.07
	الأربعاء	709	814.86	0.9203	94.54
	الخميس	721	786.00	0.9477	96.79
	الجمعة	694	766.71	0.9328	97.04

تمـــارين

(1) ما معنى تحليل السلسلة الزمنية؟ وما أغراض هذا التحليل؟

بين الأهمية العملية لتحليل السلاسل في معالجة مشكلة تنظيم مواعيد تسيير القطارات في أحدى شركات السكك الحديدية.

(2) اشرح بالتفصيل طرق استبعاد الاتجاه العام من النسب الموسمية مستعينا ببعض الأمثلة التوضيحية.

(3) الجدول التالي يبين مبيعات متجر بآلاف الدولارات في الفترة الواقعة بين 1980 وسنة 1990:

السنة	المبيعات
1980	326
1981	325
1982	331
1983	342
1984	350
1985	363
1986	359
1987	370
1988	372
1989	383
1990	388

والمطلوب:

(1) إيجاد معادلة المستقيم الذي يصف (اتجاه هذه المبيعات بطريقة المربعات الصغرى).

(2) استنتج القيم الاتجاهية للظاهرة في السنوات التالية:

1995, 1985, 1975

(3) أرصد القيم المشاهدة والقيم الاتجاهية على ورق رسم بياني.

(4) الجدول أدناه يبين كمية البنكنوت المصدر في دولة ما في آخر كل شهر من أشهر السنوات الخمس الأخيرة مقدرا بملايين الدولارات:

السنة / الشهر	1975	1976	1977	1978	1979
كانون ثاني	122	146	141	141	159
شباط	123	145	141	140	158
آذار	125	144	140	140	158
نيسان	126	143	140	139	158
أيار	127	142	139	144	155
حزيران	129	139	135	142	151
تموز	130	139	131	140	150
آب	130	139	130	137	146
أيلول	133	139	145	144	151
تشرين أول	145	147	144	155	166
تشرين ثاني	148	148	145	161	174
كانون أول	149	145	145	164	178

والمطلوب:

(1) إيجاد النسب الموسمية لحركة المبيعات.

(2) توضيح النسب المحسوبة برسم بياني يظهر الموسم الذي تزداد فيه أو تقل كمية البنكنوت المصدر.

(5) أدناه الأرقام القياسية لإنتاج السيارات في بلد صناعي في الفترة الواقعة بين سنة 1969 وسنة 1980.

السنة	1969	70	71	72	73	74	75	76	77	78	79	80
الرقم القياسي	50	58	41	66	102	91	107	108	86	110	135	85

والمطلوب:

(1) إيجاد القيم الاتجاهية للظاهرة عن طريق حساب متوسط متحرك فترته 3 سنوات وآخر فترته 4 سنوات.

(2) رصد هذه المتوسطات مع القيم الأصلية على ورق رسم بياني.

(6) توفرت البيانات التالية عن المتغيرين X ، Y:

X	1.00	1.44	1.96	3.24	4.00	7.84
Y	1	2	3	4	5	6

والمطلوب:

(1) ارسم شكل الانتشار ثم وفق خط الانحدار يدويا؟ هل ترى أن العلاقة الخطية بين المتغيرين علاقة ملائمة؟

(2) حول قيم X بأخذ الجذر التربيعي لها، ثم ارسم شكل الانتشار بين Y والمتغير الجديد $\sqrt{X}$ ؟ كيف ترى العلاقة الآن؟

(3) استخدم التحويلة أعلاه ثم أوجد خط انحدار Y على $\sqrt{X}$.

الفصل الثامن

الأرقام القياسية

(1.8) مقدمة.

(2.8) الأرقام القياسية.

(1.2.8) الأرقام التجميعية البسيطة.

(2.2.8) الأرقام التجميعية المرجحة بالأوزان.

(3.8) الأرقام القياسية النسبية.

(1.3.8) الأرقام النسبية البسيطة.

(2.3.4) الأرقام النسبية المرجحة بالأوزان.

(4.8) الأرقام القياسية ذات الأساس المتحرك.

(5.8) طرق اختيار الأرقام القياسية.

(6.8) تعديل الأرقام القياسية.

الأرقام القياسية

Index Numbers

(1.8) مقدمة (*Introduction*)

يعرف الرقم القياسي بأنه أداة لقياس التغير النسبي في قيم الظواهر من وقت لآخر أو من مكان إلى آخر. فمن المعروف أن هناك ظواهر كثيرة متغيرة تحدث في أوقات مختلفة أو أماكن مختلفة كأسعار السلع أو عدد الطلاب. وأبسط شكل يتخذه هذا الرقم هو عبارة عن نسبة مئوية تعبر عن قيمة الظاهرة في سنة معينة أو مكان معين بدلالة قيمتها في سنة أخرى تعتبر أساسا للقياس أو قيمتها في مكان آخر يعتبر أساسا للقياس أيضا. فإذا كان سعر الدقيق مثلا في سنة 1965 هو 15 دولارا للكيلو الواحد، وكان سعره في سنة 1975 هو 45 دولارا للكيلو فإن الرقم القياسي لسعر كيلو الدقيق في سنة 1975 هو:

$$\frac{45}{15} \times 100 = 300\%$$

بافتراض أن سنة 1965 أساسا للقياس. وما فعلناه هو أننا نسبنا قيمة الظاهرة في زمان معين (1975) إلى قيمتها في زمان آخر (1965)، حيث نسمي الزمن الأول بفترة المقارنة ونسمي الزمن الثاني بفترة الأساس.

وبالمقابل إذا نسبنا قيمة ظاهرة ما في مكان معين إلى قيمتها في مكان آخر نتخذه أساسا، فإننا نقول للمكان الأول مكان المقارنة ونقول للمكان الآخر مكان الأساس، وهكذا.

وبناء على هذا فانه ليس ثمة صعوبة في عمل أرقام قياسية للظواهر التي يمكن تمثيلها بقيمة واحدة في كل من السنتين المطلوب عمل المقارنة بينهما. فالرقم القياسي في هذه الحالة لا يعدو كونه نسبة مئوية بين قيمة الظاهرة في السنة المقارنة وقيمتها في السنة التي اختيرت كأساس، كالرقم القياسي لسعر كيلو الدقيق الذي تحدثنا عنه سابقا. فإذا رمزنا لسعر الدقيق في سنة المقارنة بالرمز p_n ولسعره في سنة الأساس بالرمز p_0 ، فانه يمكن التعبير عن ذلك الرقم باستخدام الرموز كما يلي:

الرقم القياسي لسعر كيلو الدقيق = $100 \times \frac{p_n}{p_0}$

وبالإضافة إلى الرقم القياسي للأسعار فانه يمكن عمل أرقام قياسية لأية ظاهرة أخرى يراد مقارنة القيم التي تتخذها بين وقت وآخر أو مكان وآخر فيمكن مثلا عمل رقم قياسي للإنتاج وآخر للصادرات وللأجور، غير أننا سنكتفي هنا بشرح طرق تركيب الأرقام القياسية للأسعار في شيء من التفصيل نظرا لأنها من أهم الأرقام القياسية وأكثرها استعمالا لتكون هذه الطرق بمثابة نماذج يمكن أن ننسج على منوالها في تركيب أي رقم قياسي آخر.

غير أن تركيب الأرقام القياسية لا يكون عادة بمثل هذه السهولة، فقد نحتاج أحيانا إلى مقارنة ظاهرتين تكون قيمة كل منها عبارة عن ملخص لعدة قيم أخرى تشترك جميعها في تكوينها. ومن الأمثلة على هذه الظواهر المركبة نفقات المعيشة (Cost of living) والمستوى العام للأسعار في سنتين مختلفتين أو في مكانين مختلفين. فمن الواضح أن مقارنة مستوى الأسعار بين مكان وآخر أو بين وقت وآخر تتطلب حساب متوسط عام يمثل جميع الأسعار في كل من المكانين أو الزمنين المطلوب عمل المقارنة بينهما. فإذا ذكرنا أن مستوى الأسعار ارتفع في سنة 1975 بمقدار 20% عنه في سنة 1965 فان هذا لا يعني طبعا أن أسعار جميع السلع في سنة 1975 تزيد بمقدار 20% عن الأسعار المناظرة لها في سنة 1965، وإنما المقصود من ذلك أن المتوسط العام للأسعار في سنة 1975 يزيد بمقدار 20% عن المتوسط العام لأسعار1965. وهذا يعني

أننا لخصنا التغيرات في أسعار جميع السلع في شكل رقم واحد يعبر عن ارتفاع أو انخفاض المستوى العام للأسعار في السنة المقارنة بالنسبة للمستوى العام للأسعار في سنة الأساس.

- **مثال (86):** الجدول التالي يمثل نفقات المعيشة لسكان حي من أحياء مدينة بغداد في السنتين 1970 & 1960، والمطلوب حساب الرقم القياسي لنفقات المعيشة باعتبار أن سنة 1960 سنة أساس.

جدول (رقم 50)

السلع	أسعار سنة 1960	أسعار سنة 1970
الخبز	7	8.1
اللحم	50.5	57.4
الخضر	8.4	8.9
منتجات الحليب	45.3	47.2
سكن، ملابس، **أخرى**	61.7	63.8
المجموع	**172.9**	**185.4**

الحل: إذا اعتبرنا أن سنة 1960 هي سنة الأساس فان:

$$\text{الرقم القياسي لتكاليف المعيشة} = \frac{\text{تكاليف المعيشة لسنة المقارنة (1970)}}{\text{تكاليف المعيشة لسنة الأساس (1960)}} (100)$$

$$= \frac{185.4}{172.9} (100) = 107\%$$

وهذا يعني أنه إذا كان متوسط تكاليف المعيشة لسكان الحي بأسعار سنة 1960 يساوي 100، فان متوسط تكاليف المعيشة بأسعار سنة 1970 لنفس سكان الحي قد ارتفع إلى ما يزيد على 107.

وتتوقف طريقة تركيب الرقم القياسي على اختيار طريقة من عدة طرق لتركيب الأرقام القياسية. غير أنه مهما اختلفت هذه الطرق (كطرق تركيب الأرقام القياسية للأسعار مثلا) **فإنها تشترك جميعها في الخطوات الآتية:**

1) اختيار السلع التي تدخل في تركيب الرقم القياسي واختيار المصادر التي تستسقى منها أسعار هذه السلع.
2) اختيار سنة الأساس: ويراعى في هذا الاختيار أن تكون سنة الأساس سنة عادية خالية من الشذوذ. فلا يصح مثلا اختيار إحدى سنوات الحرب كقاعدة للمقارنة، كما لا يجوز اختيار سنة تقع في أوج الرخاء أو في أعماق الكساد التجاري نظرا لارتفاع الأسعار أو انخفاضها بدرجة غير عادية تفسر المقارنة في مثل هذه الظروف.
3) إعطاء كل سلعة وزنا يتناسب مع أهميتها، لأن تركيب الرقم القياسي للأسعار يتطلب حساب متوسط عام لأسعار عدد كبير من السلع المختلفة الأهمية، فلابد إذن من ترجيح هذا المتوسط بالأوزان كي يكون الرقم القياسي معبرا بدقة عن حقيقة التغير في المستوى العام للأسعار.
والقاعدة العامة في اختيار الأوزان هي أن نعطي كل سلعة أهمية تتناسب مع قيمة ما يستهلك منها. وهناك طرق عدة لاختيار الأوزان تختلف باختلاف طريقة تركيب الرقم سنتحدث عنه في البنود القادمة.
4) تلخيص البيانات في شكل رقم واحد يمثل متوسط التغير النسبي في الأسعار وذلك باختيار القانون المناسب لإيجاد هذا الرقم.

وهناك طريقتان رئيسيتان لتلخيص البيانات في شكل رقم واحد يمثل متوسط التغير النسبي. ومن هاتين الطريقتين تتفرع عدة طرق تختلف باختلاف الطريقة المتبعة لاختيار الأوزان. وهاتان الطريقتان هما:

الأرقام القياسية التجميعية *Aggregate Index Numbers*

الأرقام القياسية النسبية *Relative Index Numbers*

(2.8) الأرقام القياسية التجميعية

وهي على نوعين، فهي إما أن تكون أرقاما تجميعية بسيطة أو أرقاما تجميعية مرجحة بالأوزان. وسوف نوضح كل نوع منها كما يلي.

(1.2.8) الأرقام التجميعية البسيطة للأسعار

Simple Aggregate Price Index, S.A.P.I

وفي هذه الطريقة نأخذ حاصل جمع الأسعار في السنة المقارنة كنسبة مئوية من حاصل جمع الأسعار في سنة الأساس. فإذا فرضنا أن أسعار السلع في السنة المقارنة هي p_{ni} ، وكانت أسعارها في سنة الأساس هي p_{0i} فان الرقم التجميعي البسيط للأسعار هو:

$$S.A.P.I = \frac{p_{ni}}{p_{0i}} \times 100, \qquad i = 1,2,...,r$$

حيث أن r تشير إلى عدد السلع.

$$S.A.P.I = \frac{p_{n1} + p_{n2} + ... + p_{nr}}{p_{01} + p_{02} + ... + p_{0r}} \times 100$$

$$S.A.P.I = \frac{\sum p_n}{\sum p_0} \times 100$$

- **مثال (87)**: الجدول التالي يمثل أسعارا لسلع استهلاكية معينة خلال السنتين 1960 & 1965. والمطلوب حساب الرقم التجميعي البسيط لمستوى أسعار 1965 على أساس سنة 1960.

جدول (رقم 51)

السلع	الوحدة	أسعار سنة 1960 بالدولار	أسعار سنة 1965 بالدولار
الحليب	لتر	60	75
الدقيق	كيلو	60	80
الجبن	رطل	40	45
المجموع		160	200

الحل:

هنا يكون الرقم القياسي التجميعي لمستوى الأسعار لسنة 1965 بالنسبة لسنة 1960 هو:

$$S.A.P.I = \frac{\sum p_n}{\sum p_0} \times 100$$

$$= \frac{200}{160} \times 100 = 125\%$$

وعلى الرغم من أن هذه الطريقة هي أسهل الطرق في تركيب الأرقام القياسية إلا إنها لا تستند إلى أساس صحيح. ذلك لأن أسعار السلع التي تدخل في تركيب الرقم القياسي تكون محسوبة على أساس وحدات مختلفة بعضها كبير وبعضها صغير ولذلك يكون الرقم القياسي متأثرا تأثرا عظيما بأسعار السلع ذات الوحدات الكبيرة، وهو لذلك يكون مضللا إذا كان هناك اختلاف كبير بين وحدات السلع التي تدخل في تركيبه.

(2.2.8) الأرقام التجميعية المرجحة بالأوزان

Weighted Aggregate Indexes

لكي نراعي أهمية كل سلعة عند حساب الرقم القياسي التجميعي فإننا نعطي كل سلعة وزنا يمثل الكميات المنتجة أو المستهلكة منها. وهناك أكثر من أسلوب للمفاضلة بين اختيار كميات سنة الأساس أو كميات سنة المقارنة أو المتوسط لكميات سنتي الأساس والمقارنة. **وفيما يلي عرض موجز لهذه الطرق:**

<u>أولا:</u> استخدام كميات سنة الأساس كأوزان (رقم لاسبير)

إذا فرضنا أن الكميات المستهلكة من السلع في سنة الأساس هي q_{0i} (حيث $i = 1, \ldots, r$) فان الرقم التجميعي المرجح بكميات الأساس والذي يطلق عليه رقم لاسبير التجميعي للأسعار (*Aggregate Laspeyres Price Index (A.LP.I)*) يعرف بالمعادلة التالية:

$$A.L.P.I = \frac{p_{n1}q_{01} + \ldots + p_{nr}q_{0r}}{p_{01}q_{01} + \ldots + p_{0r}q_{0r}} \times 100$$

$$= \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100$$

وبالتأمل في هذا الرقم نجد أنه متحيز إلى أعلى، ذلك لأن الكميات المستهلكة من السلع المختلفة لا تتغير بنفس النسبة شأنها شأن الأسعار التي لا تتغير بنفس النسبة. وعلى ذلك فان السلع التي ارتفعت في السعر تعطي وزنا أكبر من أهميتها الحقيقية بينما تعطي السلع التي انخفضت في السعر وزنا أقل من أهميتها الحقيقية في وقت تركيب الرقم القياسي. وحيث أننا نعطي وزنا أكبر مما يجب للسلع التي ارتفعت في الثمن ووزنا أقل مما يجب للسلع التي انخفضت في الثمن فان الرقم القياسي يكون متحيزا إلى أعلى.

ثانيا: استخدام كميات سنة المقارنة كأوزان (رقم باش)

إذا فرضنا أن كميات سنة المقارنة هي على التوالي: $q_{n1}, q_{n2}, ..., q_{nr}$ فان الرقم التجميعي المرجح بهذه الكميات (والذي يطلق عليه رقم باش التجميعي للأسعار *A.P.P.I*) تعرفه المعادلة الآتية:

$$A.P.P.I = \frac{p_{n1}q_{n1} + p_{n2}q_{n2} + ... + p_{nr}q_{nr}}{p_{01}q_{n1} + p_{02}q_{n2} + ... + p_{0r}q_{nr}} \times 100$$

$$= \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100$$

وإذا كانت الطريقة السابقة تجعل الرقم متحيزا إلى أعلى، فان هذه الطريقة تجعله متحيزا إلى أدنى، وذلك لأن السلع التي انخفضت أثمانها تزداد الكمية المستهلكة منها عادة ولذلك فهي تعطي أهمية أكثر مما يجب لمجرد أن ثمنها قد انخفض حتى كأن تغير الثمن يقرر الأوزان التي تستخدم لقياس تغير الثمن نفسه.

ثالثا: استخدام متوسط كميات سنة المقارنة وسنة الأساس كأوزان (رقم مارشال)

ويسمى برقم مارشال (*Marshal Index*)، فهو الذي اقترح استخدام متوسط كميات السنة المقارنة وسنة الأساس كأوزان وذلك للتخلص من صعوبة استخدام كميات سنة الأساس أو كميات سنة المقارنة والتي تجعل من الرقم القياسي متحيزا إلى أعلى أو متحيزا إلى أدنى على التوالي. ويعرف رقم مارشال القياسي للأسعار كما يلي:

$$M.P.I = \frac{\sum p_n (q_n + q_0)}{\sum p_0 (q_n + q_0)} \times 100$$

رابعا: الرقم القياسي التجميعي الأمثل (رقم فيشر)

وتتلخص هذه الطريقة في تركيب رقمين قياسين مستقلين يستخدم في أحدهما أوزان سنة الأساس كما فعلنا في أولا، ويستخدم في الثاني أوزان سنة المقارنة كما فعلنا

في ثانيا، والوسط الهندسي لهذين الرقمين هو الرقم القياسي الأمثل (والذي يطلق عليه رقم فيشر الأمثل *F. I. P. I*) ويعرف كالتالي:

$$F.I.P.I=\sqrt{(\frac{\sum p_n q_0}{\sum p_0 q_0})(\frac{\sum p_n q_n}{\sum p_0 q_n})}\times 100$$

- **مثال (88):** الجدول التالي يوضح أسعار وكميات سلعتين استهلاكيتين *A, B* للسنتين 1970 ,1960. المطلوب حساب الأرقام التجميعية المرجحة للأسعار باعتبار أن سنة 1960 سنة الأساس باستخدام الطرق الأربعة المشروحة في البند (2.2.8).

جدول (رقم 52)

السلعة		الأسعار		الكميات
	960	970	960	970
A	150	300	200	100
B	150	150	100	200

الحل:

(1) رقم لاسبير: ولإيجاد هذا الرقم فإننا نستخدم كميات سنة الأساس (q_0) كأوزان. والجدول التالي يبين خطوات حساب هذا الرقم على اعتبار أن سنة 1960 هي سنة الأساس.

جدول (رقم 53)

السلعة	(p)	الأسعار	(q)	الكميات	$p_n q_0$	$p_0 q_0$
	p_0	p_n	q_0	q_n		
A	150	300	200	100	60000	30000
B	150	150	100	200	15000	15000
المجموع					75000	45000

ومن الجدول نجد أن:

$$A.L.P.I = \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100$$

$$= \frac{75000}{45000} \times 100 = 167\%$$

(2) رقم باش: وفي حساب هذا الرقم نستخدم كميات سنة المقارنة (q_n) كأوزان.

والجدول التالي يبين خطوات حساب هذا الرقم باعتبار أن سنة 1960 سنة الأساس.

جدول (رقم 54)

السلعة	الأسعار (p)		الكميات (q)		$p_n q_n$	$p_0 q_n$
	p_0	p_n	q_0	q_n		
A	150	300	200	100	30000	15000
B	150	150	100	200	30000	30000
المجموع					60000	45000

ومن الحسابات أعلاه نجد أن:

$$A.P.P.I = \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100$$

$$= \frac{60000}{45000} \times 100 = 133\%$$

(3) رقم مارشال: وهنا نستخدم متوسطات كميات سنة المقارنة وكميات سنة الأساس $\left[\frac{1}{2}(q_n + q_0)\right]$ كأوزان لحساب هذا الرقم. والجدول التالي يبين خطوات الحل.

جدول (رقم 55)

السلعة	الأسعار (p)		الكميات (q)		q_n+q_0	$p_n(q_n+q_0)$	$p_0(q_n+q_0)$
	p_0	p_n	q_0	q_n			
A	150	300	200	100	300	90000	45000
B	150	100	100	200	300	30000	45000
المجموع						120000	90000

ومن الجدول فإن رقم مارشال هو:

$$M.P.I=\frac{\sum p_n(q_n+q_0)}{\sum p_0(q_n+q_0)}\times 100$$

$$=\frac{120000}{90000}\times 100=133\%$$

(4) **رقم فيشر الأمثل للأسعار:** وهو كما ذكرنا عبارة عن الوسط الهندسي لكل من رقم لاسبير ورقم باش، أي عبارة عن الجذر التربيعي لحاصل ضرب رقم لاسبير في رقم باش، أي أن

$$F.I.P.I=\sqrt{(133)(167)}=149\%$$

(3.8) الأرقام القياسية النسبية

وتكون على نوعين هي الأخرى، فهي أما أن تكون أرقاما نسبية بسيطة أو أرقاما نسبية مرجحة بالأرقام. وفيما يلي شرح لكل نوع من الأنواع.

(1.3.8) الأرقام النسبية البسيطة للأسعار

Simple Relative Price Index (S.R.P.I)

تقوم هذه الطريقة على حساب سعر كل سلعة في السنة المقارنة كنسبة مئوية من سعرها في سنة الأساس. هذه النسب المئوية تسمى مناسيب الأسعار (*Price Relative*). وبعد حساب منسوب سعر كل سلعة نعين متوسط المناسيب عن طريق

استخراج وسطها الحسابي أو الهندسي أو التوافقي ليكون هذا المتوسط هو الرقم القياسي المطلوب.

لنفرض أنه كان لدينا r من السلع، وكانت p_n هي أسعار سنة المقارنة، p_0 أسعار سنة الأساس لهذه السلع فإن الرقم القياسي النسبي البسيط للأسعار هو الوسط الحسابي للأرقام القياسية لكل سلعة على حدة ويعرف بالمعادلة الآتية:

$$S.R.P.I = (\frac{1}{r}\sum \frac{p_n}{p_0}) \times 100$$

- **مثال (89)**: في مثال (87)، احسب الرقم القياسي النسبي البسيط لأسعار 1965 بالنسبة لأسعار 1960.

الحل: يحسب الرقم القياسي النسبي البسيط كما يلي:

$$S.R.P.I = (\frac{1}{r}\sum \frac{P_n}{P_0}) \times 100$$
$$= \frac{1}{3}(\frac{75}{60} + \frac{80}{60} + \frac{45}{40})$$
$$= \left(\frac{3.708}{3}\right) \times 100 = 124\%$$

(2.3.8) الأرقام النسبية المرجحة بالأوزان

Weighted Relatives Index

ويتم حساب هذه الأرقام بإعطاء كل سلعة وزنا يتناسب مع أهميتها. وهنا يجب أن نفرق كذلك بين الأوزان التي تعتمد على سنة الأساس (رقم لاسبير) وتلك التي تعتمد على سنة المقارنة (رقم باش). وفيما يلي توضيح لكل نوع من هذه الأرقام.

أولا: رقم لاسبير النسبي للأسعار

Relative Laspeyres Price Index (R.L.P.I)

وهو يعتمد على أوزان سنة الأساس كما أسلفنا. فإذا كانت q_0, p_0 هي أسعار وكميات سنة الأساس، وكانت p_n هي أسعار سنة المقارنة فإن رقم لاسبير النسبي للأسعار لسنة المقارنة بالنسبة لسنة الأساس هو:

$$R.L.P.I = \frac{\sum (\frac{p_n}{p_0}) p_0 q_0}{\sum p_0 q_0} \times 100$$

وإذا وضعنا $w_0 = \frac{p_0 q_0}{\sum p_0 q_0}$ وهي تمثل أوزان سنة الأساس، فإن الرقم القياسي أعلاه يصبح بالشكل التالي:

$$R.L.P.I = (\sum \frac{p_n}{p_0} \times w_0) \times 100$$

ثانيا: رقم باش النسبي للأسعار *(Relative Paasche Price Index (R.P.P.I*

ويعتمد هذا الرقم في حسابه على أوزان سنة المقارنة. فإذا افترضنا أن q_n, p_n هي أسعار وكميات سنة المقارنة، وأن p_0 هي أسعار سنة الأساس فإن رقم باش النسبي للأسعار لسنة المقارنة بالنسبة لسنة الأساس هو:

$$R.P.P.I = \frac{\sum (\frac{p_n}{p_0}) p_n q_n}{\sum p_n q_n} \times 100$$

فإذا فرضنا أن $w_n = \frac{p_n q_n}{\sum p_n q_n}$ وهي تمثل أوزان سنة الأساس، فإن معادلة الرقم أعلاه تكون على الشكل التالي:

$$R.P.P.I = (\sum \frac{p_n}{p_0} \times w_n) \times 100$$

ثالثا: رقم فيشر الأمثل النسبي للأسعار

وهو عبارة عن الوسط الهندسي لكل من رقم لاسبير ورقم باش النسبيين للأسعار، أي أن:

$$F.R.P.I=\sqrt{(\sum\frac{p_n}{p_0}\times w_0)(\sum\frac{p_n}{p_0}\times w_n)}$$

- **مثال (90):** للبيانات الواردة في مثال (88)، المطلوب حساب الأرقام القياسية النسبية المرجحة للأسعار بالطرق الثلاثة السابقة باعتبار أن سنة 1960 سنة الأساس.

الحل:

(1) رقم لاسبير النسبي المرجح للأسعار: الجدول الآتي يبين حساب هذا الرقم.

جدول (رقم 56)

السلعة	الأسعار (p)		الكميات (q)		p_0q_0	w_0
	p_0	p_n	q_0	q_n		
A	150	300	200	100	30000	$\frac{30000}{45000}=0.67$
B	150	150	100	200	15000	$\frac{15000}{45000}=0.33$
المجموع					45000	

وبالتالي فإن رقم لاسبير النسبي المرجح للأسعار هو:

$$R.L.P.I=(\sum\frac{p_n}{p_0}\times w_0)\times 100$$
$$=(\frac{300}{150})(0.67)+(\frac{150}{150})(0.33)$$
$$=1.67\times 100=167\%$$

(2) رقم باش النسبي المرجح للأسعار: الجدول التالي يبين خطوات طريقة الحساب.

جدول (رقم 57)

السلعة	(p)	الأسعار	(q)	الكميات	$p_n q_n$	w_n
	p_0	p_n	q_0	q_n		
A	150	300	200	100	30000	$\frac{30000}{60000}=0.5$
B	150	150	100	200	30000	$\frac{30000}{60000}=0.5$
المجموع					60000	

ومن الجدول نجد أن رقم باش النسبي للأسعار يساوي:

$$R.P.P.I=\left(\sum \frac{P_n}{P_0}\times w_n\right)\times 100$$

$$=\left(\frac{300}{150}\right)(0.5)+\left(\frac{150}{150}\right)(0.5)$$

$$=1.50\times 100=150\%$$

(3) رقم فيشر النسبي للأسعار: وهو عبارة عن الجذر التربيعي لحاصل ضرب كل من رقم لاسبير وباش النسبيين المرجحين للأسعار، أي أن:

$$R.F.P.I=\sqrt{(150)(167)}=158\%$$

(4.8) الأرقام القياسية ذات الأساس المتحرك

Index Numbers with Moving Base

تعتبر سنة الأساس موضوع على جانب من الأهمية عند تركيب الرقم القياسي. وفي دراستنا السابقة كنا نعتبر أن الأساس ثابت، غير أنه بابتعاد سنة الأساس كثيرا عن السنة المقارنة نجد أن هناك بعض السلع الهامة التي لم تكن موجودة في سنة الأساس أو على الأقل لم تكن موجودة بشكلها الحالي. فإذا أردنا مثلا عمل رقم قياسي لتكاليف

النقل من الرياض إلى جدة فلا يصح مثلا اتخاذ سنة 1900 كسنة أساس، إذ أن وسائل النقل في تلك السنة تختلف اختلافا كليا عن وسائل النقل في الوقت الحاضر. والحال كذلك عند حسابنا للرقم القياسي لتكاليف المعيشة مثلا، فنحن سوف نجد سلعا جديدة تدخل إلى السوق وسلعا أخرى تختفي منه. لذلك نحن نلجأ إلى تغيير سنة الأساس من وقت لآخر. والسبب في تغيير سنة الأساس هو أنه لكي تكون المقارنة صحيحة فلابد أن تكون الظروف المحيطة بإنتاج واستهلاك السلع في السنة المقارنة لا تختلف كثيرا عن الظروف في سنة الأساس ولكي نتغلب على صعوبة تغير الظروف الحالية عن الظروف المناظرة لها في سنة الأساس نستخدم طريقة الأساس المتحرك. وفي هذه الحالة تحسب أسعار كل سنة كنسبة مئوية من أسعار السنة السابقة لها مباشرة، ثم وبضرب تلك الأرقام في بعضها نحصل على الرقم القياسي المطلوب.

- **مثال (91)**: الجدول التالي يبين أسعار السلعتين A, B من سنة 1975 وحتى سنة 1984 **والمطلوب**:

(أ) إيجاد الرقم القياسي النسبي للأسعار لسنة 1984 بالنسبة لأسعار سنة 1975.

(ب) إيجاد الرقم القياسي للأسعار لسنة 1984 بالنسبة لأسعار سنة 1975 بطريقة الأساس المتحرك.

(ج) باعتبار أن سنة الأساس الجديدة هي سنة 1981، المطلوب إيجاد الأرقام القياسية للأسعار من سنة 1975 وحتى سنة 1984.

جدول (رقم 58)

السلعة	الأسعار									
	75	76	77	78	79	80	81	82	83	84
A	8	8	8	10	10	13	17	28	33	38
B	13	13	13	13	18	18	19	19	23	23

الحل:

(أ) الرقم القياسي النسبي لأسعار سنة 1984 بالنسبة لأسعار سنة 1975 على الأساس الثابت هو:

$$S.R.P.I = \frac{1}{r}\sum \frac{p_n}{p_0}$$
$$= \frac{1}{2}\left(\frac{38}{8} + \frac{23}{13}\right) = (3.26) \times 100$$
$$= 326\%$$

(ب) ما سوف نقوم به الآن هو تكوين عدة أرقام قياسية لكل فترة على حدة باتخاذ الفترة السابقة لها كأساس للمقارنة. بمعنى أننا نحسب الرقم القياسي النسبي لأسعار 76 بالنسبة لأسعار 75، ثم الرقم القياسي النسبي لأسعار 77 بالنسبة لعام 76 وهكذا حتى نصل إلى الرقم القياسي النسبي لأسعار 84 بالنسبة لأسعار 83، وبضرب هذه الأرقام القياسية بعضها بالبعض الآخر نحصل على الرقم القياسي النسبي لأسعار 84 بالنسبة لأسعار 75 على الأساس المتحرك. وبالتالي فإن الرقم القياسي النسبي لأسعار 76 بالنسبة لأسعار 75 هو:

$$S.R.P.I(1976/75) = \frac{1}{2}(8/8 + 13/13) = 1$$

وتكون بقيمة الأرقام القياسية الأخرى كما يلي:

$$S.R.P.I(77/76) = 1.00, \quad S.R.P.I(78/77) = 1.13, \quad S.R.P.I(79/78) = 1.19$$
$$S.R.P.I(80/79) = 1.15, \quad S.R.P.I(81/80) = 1.18, \quad S.R.P.I(82/81) = 1.32$$
$$S.R.P.I(83/82) = 1.19, \quad S.R.P.I(84/83) = 1.08.$$

وبذلك يكون الرقم القياسي النسبي لأسعار 84 بالنسبة لأسعار 75 هو:

$$S.R.P.I(1975/1984) = 1 \times 1 \times 1.13 \times 1.19 \times 1.15 \times 1.18 \times 1.32 \times 1.19 \times 1.08$$
$$= (3.08) \times 100 = 308\%$$

(ج) إذا اعتبرنا أن سنة 1981 هي فترة الأساس الجديدة فإن هناك طريقين لإيجاد الأرقام القياسية الجديدة:

الأول: إعادة حساب جميع الأرقام القياسية مستعملين السنة الجديدة كأساس.

الثاني: نقسم جميع الأرقام القياسية للسنوات المختلفة على الرقم القياسي المقابل لسنة الأساس الجديدة.

وهذا ما سنقوم به في هذا المثال، حيث يمثل الصف الأول من الجدول أدناه الأرقام القياسية السابقة على اعتبار أن سنة الأساس هي السنة 1975، ويمثل الصف الثاني الأرقام القياسية الجديدة على اعتبار أن سنة الأساس هي السنة 1981 كما هو المطلوب في السؤال.

جدول (رقم 59)

السنة	75	76	77	78	79	80	81	82	83	84
الرقم القديم	1	1	1	1.13	1.19	1.15	1.18	1.32	1.19	1.08
الرقم الجديد	0.85	0.85	0.85	0.95	1.01	0.97	1	1.12	1.01	0.92

(5.8) طرق اختبار الأرقام القياسية

Tests of Index Numbers Methods

رأينا في دراستنا السابقة أن هناك عددا كبيرا من الطرق المختلفة لتركيب الأرقام القياسية. وقد لاحظنا أن كل طريقة من هذه الطرق تعطينا نتيجة تختلف عن تلك التي نحصل عليها بأية طريقة أخرى. ولكي نتأكد من دقة النتائج التي نحصل عليها، لابد من أن نبحث عن وسيلة للمفاضلة بين هذه الطرق المختلفة. وفي هذا البند سنتعرض إلى بعض الاختبارات التي سنطبقها على الصيغ المختلفة لتركيب الأرقام القياسية. **ومن هذه الاختبارات المقترح تطبيقها هي:**

- اختبار الانعكاس الزمني *Time Reversal Test*
- اختبار الانعكاس المعاملي *Factor Reversal Test*

أولا: اختبار الانعكاس الزمني

ولتطبيق هذا الاختبار نضرب كل رقم قياسي في بديله الزمني (*Time Reciprocal*) والذي نحصل عليه من تبديل وضع سنة الأساس وسنة المقارنة لنرى ما إذا كان حاصل الضرب يساوي الواحد الصحيح. وبعبارة أخرى إذا عكسنا وضع سنة الأساس وسنة المقارنة واستخرجنا رقما قياسيا جديدا فلابد أن يكون حاصل ضرب الرقم الأصلي في الرقم الجديد يساوي واحد.

هذا هو اختبار الانعكاس الزمني . وعلى الرغم من أن تحقق شرط الانعكاس الزمني يبدو كبديهة واضحة لا تحتاج إلى اختبار، إلا أن كثيرا من الصيغ المذكورة آنفا لا تحقق هذا الشرط كما سنرى لاحقا. وفيما يلي سنتناول بعض الصيغ السالفة لنرى أيها يحقق اختبار الانعكاس في الزمن.

(1) الرقم التجميعي البسيط

إذا قمنا بإهمال ضرب الصيغ المختلفة للأرقام القياسية في 100 لتسهيل الكتابة، فإن الرقم التجميعي البسيط للأسعار هو $\sum p_n / \sum p_0$ حيث: p_n تمثل أسعار سنة المقارنة، p_0 تمثل أسعار سنة الأساس وأن بديله الزمني هو $\sum p_0 / \sum p_n$ ، وحيث أن:

$$\frac{\sum p_n}{\sum p_0} \times \frac{\sum p_0}{\sum p_n} = 1$$

لذا فإن هذا الرقم يحقق شرط الانعكاس في الزمن. لاحظ أن تحقيق هذا الرقم لشرط الانعكاس في الزمن لا يجعل منه رقما كاملا، فقد ذكرنا قبلا أن هذا الرقم لا يصلح مطلقا إذا كان هناك تفاوت كبير بين وحدات السلع الداخلة في تركيبه.

(2) الأرقام التجميعية المرجحة بالأوزان

(أ) رقم لاسبير التجميعي للأسعار: إذا تناولنا هذا الرقم الذي يرجح بأوزان سنة الأساس والذي يعرف بالآتي:

$$A.L.P.I=\frac{\sum p_n q_0}{\sum p_0 q_0}$$

فإن بديله الزمني هو: $\frac{\sum p_0 q_n}{\sum p_n q_n}$

ولأن رقم لاسبير في بديله الزمني لا يساوي الواحد الصحيح، أي أن

$$\frac{\sum p_n q_0}{\sum p_0 q_0}\times\frac{\sum p_0 q_n}{\sum p_n q_n}\neq 1$$

لذلك فإن هذا الرقم لا يحقق شرط الانعكاس في الزمن. وبنفس الطريقة يمكن إثبات أن الرقم التجميعي المرجح بأوزان السنة المقارنة (رقم باش) لا يحقق هذا الشرط.

(ب) رقم مارشال للأسعار: وهو الرقم التجميعي المرجح بمتوسط أوزان سنة الأساس والسنة المقارنة، ويعرف كالتالي:

$$A.M.P.I=\frac{\sum p_n(q_n+q_0)}{\sum p_0(q_n+q_0)}$$

وحيث أن:

$$\frac{\sum p_n(q_0+q_n)}{\sum p_0(q_0+q_n)}\times\frac{\sum p_0(q_n+q_0)}{\sum p_n(q_n+q_0)}=1$$

فإن الرقم يحقق شرط الانعكاس الزمني.

(ج) **الرقم القياسي الأمثل:** وهو عبارة عن الجذر التربيعي لرقم لاسبير في رقم باش، أي أن:

$$A.F.P.I=\sqrt{\frac{\sum p_nq_0}{\sum p_0q_0}\times\frac{\sum p_nq_n}{\sum p_0q_n}}$$

وحيث أن:

$$\sqrt{(\frac{\sum p_nq_0}{\sum p_0q_0}\times\frac{\sum p_nq_n}{\sum p_0q_n})(\frac{\sum p_0q_n}{\sum p_nq_n}\times\frac{\sum p_0q_0}{\sum p_nq_0})}=1$$

لذلك فإن الرقم القياسي الأمثل يحقق شرط الانعكاس في الزمن.

ثانيا: اختبار الانعكاس المعاملي

إذا حسبنا رقما قياسيا لأسعار بعض السلع باستعمال صيغة معينة ثم حسبنا رقما قياسيا لكميات هذه السلع باستعمال نفس الصيغة فلابد أن يكون حاصل ضرب الرقمين يساوي الرقم القياسي للقيمة، أي:

الرقم القياسي للأسعار × الرقم القياسي للكميات = الرقم القياسي للقيمة

وهذا هو اختبار الانعكاس المعاملي. وللحصول على الرقم القياسي للكميات بنفس الطريقة التي حصلنا بها على الرقم القياسي للأسعار. فإننا نعود إلى الصيغة التي حسبنا بها الرقم القياسي للأسعار ونبدل فيها كل (p) بـ (q) وكل (q) بـ (p). والصيغة الجديدة بعد هذا التبديل تسمى البديل المعاملي (*Factor Reciprocal*). ولتحقيق شرط الانعكاس المعاملي لابد أن يكون:

الرقم القياسي × بديله المعاملي = الرقم القياسي للقيمة

وفيما يلي تطبيق اختبار الانعكاس المعاملي على بعض الأرقام القياسية.

(1) الرقم التجميعي البسيط للأسعار:

$$S.A.P.I=\frac{\sum p_n}{\sum p_0}$$

وإن بديله المعاملي هو $\frac{\sum q_n}{\sum q_0}$ وهو الرقم التجميعي البسيط للكميات.

وحيث أن: $$\frac{\sum p_n}{\sum p_0}\times\frac{\sum q_n}{\sum q_0}\neq\frac{\sum p_n q_n}{\sum p_0 q_0}$$

نجد أن هذا الرقم لا يحقق شرط الانعكاس المعاملي.

(2) رقم لاسبير التجميعي للأسعار:

$$A.L.P.I=\frac{\sum p_n q_0}{\sum p_0 q_0}$$

لذا فإن بديله المعاملي هو: $\frac{\sum q_n p_0}{\sum q_0 p_0}$ وهو رقم لاسبير التجميعي للكميات. ومن الواضح أن حاصل ضرب هذين الرقمين لا يساوي الرقم القياسي للقيمة. مما يعني أن هذا الرقم لا يجتاز هذا الاختبار. وبتطبيق هذا الاختبار على الرقم التجميعي المرجح بأوزان المقارنة (رقم باش) نجد انه هو الآخر لا ينعكس في المعامل.

(3) الرقم القياسي الأمثل للأسعار:

$$A.F.P.I=\sqrt{\frac{\sum p_n q_0}{\sum p_0 q_0}\times\frac{\sum p_n q_n}{\sum p_0 q_n}}$$

لذا فإن بديله المعاملي هو $\sqrt{\frac{\sum q_n p_0}{\sum q_0 p_0}\times\frac{\sum q_n p_n}{\sum q_0 p_n}}$ وهو رقم فيشر الأمثل للكميات.

ومن الواضح أن حاصل ضرب هذين الرقمين

$$= \sqrt{\left(\frac{\sum p_n q_n}{\sum p_0 q_0}\right)^2} = \frac{\sum p_n q_n}{\sum p_0 q_0}$$

ومن هذا يتضح أن الرقم القياسي الأمثل يحقق شرط الانعكاس المعاملي علاوة على تحقيقه لشرط الانعكاس الزمني. ومن هنا جاءت تسميته بالرقم الأمثل.

- **مثال (92)**: للسلعتين A,B المذكورتين في مثال (88)، المطلوب تطبيق اختباري الانعكاس في الزمن والانعكاس في المعامل على بعض الأرقام القياسية.

الحل: أولاً: اختبارات الانعكاس في الزمن

(أ) الرقم القياسي التجميعي البسيط للأسعار: لنتذكر أسعار وكميات السلعتين A, B الواردة في جدول (رقم 52).

جدول (رقم 60)

السلعة	الأسعار P		الكميات q	
	p_0	p_n	q_0	q_n
A	150	300	200	100
B	150	150	100	200
المجموع	300	450	300	300

ومن الجدول نجد أن الرقم القياسي التجميعي البسيط لأسعار 1970 بالنسبة لأسعار 1960 هو:

$$S.A.P.I = \frac{\sum p_n}{\sum p_0} = \frac{450}{300}$$

وأن الرقم القياسي التجميعي البسيط لأسعار 1960 بالنسبة لأسعار 1970 يساوي $\frac{300}{450}$ وهذان الرقمان القياسيان متبادلان حاصل ضربهما يساوي واحد، أي أن:

$$\frac{450}{300} \times \frac{300}{450} = 1$$

مما يعني أن هذا الرقم يحقق اختبار الانعكاس في الزمن أو الأساس.

(ب) **الرقم القياسي النسبي البسيط للأسعار:** إذا كانت سنة 1960 هي سنة الأساس، فإن:

$$S.R.P.I = \frac{1}{r}\sum(\frac{p_n}{p_0})$$
$$= \frac{1}{2}(\frac{300}{150} + \frac{150}{150}) = \frac{3}{2}$$

وأن البديل الزمني هو: $\frac{1}{2}(\frac{150}{300} + \frac{150}{150}) = \frac{1.5}{2}$ باعتبار أن سنة 1970 هي سنة الأساس. وهذان الرقمان غير متبادلين، بمعنى أن حاصل ضربهما لا يساوي الواحد الصحيح، وذلك لأن:

$$\frac{3}{2} \times \frac{1.5}{2} \neq 1$$

وهذا يؤكد أن الرقم القياسي النسبي البسيط لا يحقق اختبار الانعكاس في الزمن. لاحظ أن حاصل ضرب الرقم القياسي النسبي البسيط في بديله الزمني أكبر من الواحد الصحيح، وهنا يقال أن هذا الرقم منحاز للأعلى. أما إذا كان حاصل ضرب الرقم القياسي في بديله الزمني أقل من واحد فيقال أن الرقم منحاز للأدنى.

(ج) **رقم لاسبير التجميعي للأسعار:** إذا اعتبرنا أن سنة 1960 هي سنة الأساس فإن:

$$A.L.P.I = \frac{\sum p_n q_0}{\sum p_0 q_0} = \frac{75000}{45000} = 1.67$$

انظر جدول (رقم 53). أما الجدول التالي فهو يبين البيانات اللازمة لحساب البديل الزمني للرقم السابق باعتبار أن سنة 1970 هي سنة الأساس.

جدول (رقم 61)

السلعة	(p) الأسعار 1960 p_0	1970 p_n	(q) الكميات 1960 q_0	1970 q_n	p_0q_n	p_nq_n	p_0q_0	p_nq_0
A	150	300	200	100	15000	30000	30000	60000
B	150	150	100	200	30000	30000	15000	15000
المجموع					45000	60000	45000	75000

ومن الجدول نجد أن البديل الزمني $= \frac{\sum p_0q_n}{\sum p_nq_n} = \frac{45000}{60000}$ ، وحيث أن:

$$\frac{75000}{45000} \times \frac{45000}{60000} \neq 1$$

فإن رقم لاسبير التجميعي لا يجتاز اختبار الانعكاس في الزمن، وهو متحيز إلى أعلى طالما أن نتيجة حاصل الضرب اكبر من واحد. وهذا ما كنا قد ذكرناه سابقا.

(د) رقم فيشر الأمثل:

$$A.F.P.I = \sqrt{(\frac{\sum p_nq_0}{\sum p_0q_0})(\frac{\sum p_nq_n}{\sum p_0q_n})}$$

وأن بديله الزمني هو:

$$\sqrt{(\frac{\sum p_0q_n}{\sum p_nq_n})(\frac{\sum p_0q_0}{\sum p_nq_0})}$$

فإنه وباستخدام البيانات المذكورة في الجداول المرقمة (53 , 54 , 61) نجد أن رقم فيشر الأمثل في بديله الزمني يساوي الآتي:

$$\sqrt{(\frac{75000}{45000})(\frac{60000}{45000}) \times (\frac{45000}{60000})(\frac{45000}{75000})} = 1$$

مما يعني أن الرقم القياسي الأمثل للأسعار يجتاز اختبار الانعكاس في الزمن.

ثانيا: اختبارات الانعكاس في المعامل

(أ) الرقم التجميعي البسيط للأسعار: بالإضافة إلى أعمدة الجدول (رقم 60)، سوف نضيف عمودا آخر بقيم عام 1960 وثاني بقيم عام 1970.

جدول (رقم 62)

السلعة	*P* (1960) p_0	الأسعار (1970) p_n	*Q* (1960) q_0	الكميات (1970) q_n	*V* (1960) V_0	القيم (1970) V_n
A	150	300	200	100	30000	30000
B	150	150	100	200	15000	30000
المجموع	300	450	300	300	45000	60000

ومن الجدول السابق نجد أن الرقم القياسي التجميعي البسيط للأسعار هو:

$$S.A.P.I=\frac{\sum p_n}{\sum p_0}=\frac{450}{300}$$

والرقم القياسي التجميعي البسيط للكميات هو:

$$S.A.Q.I=\frac{\sum q_n}{\sum q_0}=\frac{300}{300}$$

وأخيرا فان الرقم القياسي التجميعي البسيط للقيم هو:

$$S.A.V.I=\frac{\sum p_n q_n}{\sum p_0 q_0}=\frac{\sum V_n}{\sum V_0}=\frac{60000}{45000}$$

ولكن:

$$(\frac{450}{300})(\frac{300}{300})\neq\frac{60000}{45000}$$

لذلك فالرقم التجميعي البسيط للأسعار لا يحقق اختبار الانعكاس ألمعاملي.

(ب) الرقم القياسي النسبي البسيط للأسعار

الرقم القياسي النسبي البسيط للأسعار هو 3/2 كما وجدناه في الفقرة السابقة.

أما الرقم القياسي النسبي البسيط للكميات (انظر جدول 62) فهو:

$$S.R.Q.I = \frac{1}{r}\sum\frac{q_n}{q_0}$$
$$= \frac{1}{2}(\frac{100}{200}+\frac{200}{100}) = \frac{2.5}{2}$$

والرقم القياسي النسبي البسيط للقيم يساوي:

$$S.R.V.I = \frac{1}{r}\sum\frac{V_n}{V_0}$$
$$= \frac{1}{2}(\frac{30000}{30000}+\frac{30000}{15000}) = \frac{3}{2}$$

وحيث أن:

$$\frac{3}{2}\times\frac{2.5}{2}\neq\frac{3}{2}$$

فان الرقم النسبي البسيط للأسعار هو الآخر لا يحقق اختبار الانعكاس في المعامل.

(ج) الرقم القياسي الأمثل للأسعار

$$A.F.P.I = \sqrt{(\frac{\sum p_n q_0}{\sum p_0 q_0})(\frac{\sum p_n q_n}{\sum p_0 q_n})}$$

فان بديله المعاملي هو:

$$\sqrt{(\frac{\sum q_n p_0}{\sum q_0 p_0})(\frac{\sum q_n p_n}{\sum q_0 p_n})}$$

وباستخدام الأرقام والبيانات الواردة في الجداول (53 ,54 ,62) نجد أن حاصل ضرب الرقم القياسي الأمثل في بديله المعاملي يساوي الرقم القياسي للقيمة كما سنرى فيما يأتي.

$$= \sqrt{(\frac{75000}{45000})(\frac{60000}{45000})(\frac{45000}{45000})(\frac{60000}{75000})}$$
$$= \sqrt{(\frac{60000}{45000})^2} = \frac{60000}{45000} = \frac{4}{3}$$

مما يعني أن الرقم القياسي الأمثل يجتاز مرة أخرى اختبار الانعكاس في المعامل.

(6.8) تعديل الأرقام القياسية

رأينا أن هناك صيغا كثيرة لا تجتاز اختبارات الأرقام القياسية، ولكنه بالإمكان تعديل أي صيغة تستعمل لتركيب الرقم القياسي لجعلها تنعكس في الزمان والمعامل.

(أ) تعديل الرقم لجعله يحقق شرط الانعكاس الزمني

إذا أردنا تعديل أي رقم قياسي لجعله يحقق شرط الانعكاس الزمني فإننا نضرب هذا الرقم في مقلوب بديله الزمني ثم نستخرج الجذر ألتربيعي لحاصل الضرب. هذا يعني أن الرقم القياسي المعدل يساوي الوسط الهندسي للرقم الأصلي ومقلوب بديله الزمني. والمثال التالي يوضح هذه الطريقة.

- **مثال (93):** المطلوب إجراء التعديل المذكور على الرقم التجميعي المرجح بأوزان سنة سنة.

الحل:

نحن نعلم أن رقم لاسبير للأسعار هو من الأرقام التي لا تنعكس في الزمن. ويتم إجراء التعديل المذكور عليه كما يلي:

الرقم الأصلي هو: $\frac{\sum p_0 q_n}{\sum p_n q_n}$ وان بديله الزمني هو: $\frac{\sum p_n q_0}{\sum p_0 q_0}$ وبالتالي فان المقلوب للبديل الزمني هو: $\frac{\sum p_n q_n}{\sum p_0 q_n}$. أما الرقم المعدل فهو عبارة عن الجذر ألتربيعي لحاصل ضرب الرقم الأصلي في مقلوب بديله الزمني، أي أن:

$$\sqrt{(\frac{\sum p_n q_0}{\sum p_0 q_0})(\frac{\sum p_n q_n}{\sum p_0 q_n})}$$

وهذا الرقم المعدل هو نفس الرقم القياسي الأمثل في الزمن كما رأينا.

(ب) تعديل الرقم لجعله يحقق شرط الانعكاس ألمعاملي

إذا أردنا تعديل الأرقام القياسية لجعلها تحقق شرط الانعكاس المعاملي فإننا نحسب الوسط الهندسي للرقم الأصلي ومقلوب بديله ألمعاملي،كما فعلنا في الحالة السابقة علما بأن:

المقلوب المعاملي = الرقم القياسي للقيمة / البديل المعاملي للرقم المطلوب تعديله

- **مثال(94):** المطلوب عمل التعديل اللازم لجعل الرقم التجميعي المرجح بكميات سنة الأساس ينعكس في المعامل.

الحل:

نحن نعلم أن الرقم أعلاه وهو رقم لاسبير لا ينعكس في المعامل، ولجعله كذلك فان:

الرقم الأصلي هو: $\frac{\sum p_n q_0}{\sum p_0 q_0}$ وان بديله الزمني هو: $\frac{\sum q_n p_0}{\sum q_0 p_0}$ وبالتالي فان مقلوبه المعاملي يكون:

$$= \frac{\sum p_n q_n}{\sum p_0 q_0} \div \frac{\sum q_n p_0}{\sum q_0 p_0}$$

$$= \frac{\sum p_n q_n}{\sum p_0 q_n}$$

إذن الوسط الهندسي للرقم الأصلي ومقلوبه المعاملي:

$$\sqrt{\left(\frac{\sum p_n q_0}{\sum p_0 q_0}\right)\left(\frac{\sum p_n q_n}{\sum p_0 q_n}\right)}$$

وهذا هو الرقم القياسي الأمثل والذي ينعكس في المعامل كما أثبتنا.

وفي الحقيقة إذا كان الرقم القياسي لا ينعكس في الزمن ولا في المعامل وجب تعديله مرتين، حيث نعدله أولا لنجعله يحقق شرط الانعكاس في الزمن ثم نتناول الصيغة المعدلة فنعدلها مرة أخرى لنجعلها تحقق شرط الانعكاس في المعامل.

غير أن الصيغة النهائية المعدلة بهذه الطريقة كثيرا ما تكون بالغة التعقيد بحيث يصعب استعمالها. ناهيك عن أنها تأخذنا بعيدا عن الغرض الأصلي الذي نرمي إليه من تركيب الرقم القياسي وهو قياس متوسط التغير النسبي في قيم الظواهر. ويبقى الرقم القياسي الأمثل هو أفضل الأرقام القياسية السابقة فهو يحقق شرطي الانعكاس الزمني و المعاملي معا، فضلا عن أن صيغته بسيطة و خالية من التعقيد.

تمـــارين

(1) الجدول التالي يمثل أسعار وكميات ثلاثة سلع لعامي 1983 ، 1985.

السلعة	الأسعار		الكميات	
	1983	**1985**	**1983**	**1985**
A	80	82	80	90
B	83	85	110	100
C	82	84	90	80

والمطلوب حساب الأرقام القياسية أدناه معتبرا أن عام 1983 سنة الأساس.

(1) الرقم التجميعي البسيط للأسعار.

(2) الرقم النسبي البسيط للأسعار.

(3) رقم لاسبير النسبي والتجميعي للأسعار.

(4) رقم باش النسبي والتجميعي للأسعار.

(5) رقم فيشر الأمثل النسبي والتجميعي للأسعار.

(6) رقم مارشال للكميات المرجح بالأسعار.

(7) الرقم التجميعي البسيط للكميات.

(8) الرقم النسبي البسيط للكميات.

(2) اشرح طريقة تعديل الأرقام القياسية النسبية والتجميعية لجعلها تحقق شرطي الانعكاس الزمني والانعكاس المعاملي.

(3) اشرح أهمية الترجيح بالأوزان عند حساب الأرقام القياسية.

(4) المصنعان A , B متخصصان بإنتاج سلعة معينة وكانت كميات الإنتاج وسعر البيع من المصنع لهذه المادة في سنتي 1986 ، 1987 يبينها الجدول التالي:

المصنع	الأسعار		كميات	الإنتاج
	1986	**1987**	**1986**	**1987**
A	60	62	80	70
B	65	67	60	50

المطلوب: (1) حساب جميع الأرقام النسبية والتجميعية البسيطة والمرجحة للأسعار.

(2) معرفة أي الأرقام في أعلاه:

منحاز للأعلى.

منحاز للأدنى.

غير منحاز.

(3) أي من الأرقام القياسية في المطلوب الأول تحقق الانعكاس في المعامل؟ اثبت ذلك؟

قائمة المراجع

أولاً: المراجع العربية:

1) **أبو صالح، محمد صبحي و عوض، عدنان محمد** (1983)، مقدمة في الإحصاء. نيويورك، دار جون وايلي وأبنائه.

2) **الراوي، خاشع محمود** (1984)، المدخل إلى الإحصاء. الموصل: جامعة الموصل، العراق.

3) **الهانسي، مختار محمود** (1984)، مقدمة في طرق التحليل الإحصائي. دار النهضة العربية للطباعة والنشر، بيروت، لبنان.

4) **البدران، غضبان عبد الله خلف** (2003)، طرق التحليل الإحصائي. مركز عبادي للدراسات والنشر، صنعاء، اليمن.

5) بري، عدنان بن ماجد و هندي، محمود محمد و عبد الله، أنور أحمد (1977)، مبادئ الإحصاء والاحتمالات. جامعة الملك سعود، الرياض، المملكة العربية السعودية.

6) **عقيل، محمد بن إبراهيم و أبوعمة، عبد الرحمن بن محمد** (2000)، نظرية الاحتمالات وتطبيقاتها. جامعة الملك سعود، الرياض، المملكة العربية السعودية.

7) **عاشور، سمير كامل** (1980)، مقدمة في الإحصاء التحليلي. معهد الدراسات والبحوث الإحصائية، جامعة القاهرة: القاهرة، مصر.

8) **كاظم، أموري هادي و الدليمي، محمد مناجد** (1988)، مقدمة في تحليل الانحدار الخطي. جامعة بغداد: بغداد، العراق.

ثانياً: المراجع الأجنبية:

1) Raymond S. Nickerson(2003). **Psychology and Environmental Change**, Britain: Psychology Press.
2) **Draper, N. and Smith, H**. (1981), Applied Regression Analysis. John Wiley and Sons, Inc.
3) **Dunn, G. and Clark, J**. (1974), Applied Statistics . John Wiley and Sons, Inc.
4) **Harnett, D. L**. (1975), Introduction to Statistical Methods, 2nd edition, Addison – Wesley Publishing Com.
5) **Hoel, P. G. (1970)**, Elementary of Statistics, 4th edition, John Wiley and Sons, Inc.
6) **Hoel, P. G. and Jessen, A**. (1977), Basic Statistics for Business and Economics, 2nd edition, John Wiley and Sons, Inc.
7) **Koutsoyiannis, A**. (1977), Theory of Econometrics, 2nd edition, Macmillan Company.
8) **Sahai, H. and Ageel, M, I**. (2000), The Analysis of Variance, Birkhauser Boston.
9) **Seber, G. A. F.** (1977), Linear Regression Analysis, John Wiley and Sons, Inc.
10) **Theirl, H.** (1971), Principle of Econometric, John Wiley and Sons, Inc.

الجداول الإحصائية

أولا : جدول الأرقام العشوائية.

ثانيا: جدول التوزيع الطبيعي المعياري.

ثالثا: جدول توزيع كاي تربيع.

رابعا: جدول توزيع t .

خامسا: جدول توزيع F .

سادسا: جدول تحويل r **إلى** Z.

سابعا: جدول كولومجروف.

جدول الأرقام العشوائية

08	35	86	99	10	78	54	24	27	85	13	66	15	88	73	04	61	89	75	53
28	30	60	32	64	81	33	31	05	91	40	51	00	78	93	32	60	46	04	75
53	84	08	62	33	81	59	41	36	28	51	21	59	02	90	28	46	66	87	95
91	75	75	37	41	61	61	36	22	69	50	26	39	02	12	55	78	17	65	14
89	41	59	26	94	00	39	75	83	91	12	60	71	76	46	48	94	97	23	06
77	51	30	38	20	86	83	42	99	01	68	41	48	27	74	51	90	81	39	80
19	50	23	71	74	69	97	92	02	88	55	21	02	97	73	74	28	77	52	51
21	81	85	93	13	93	27	88	17	57	05	68	67	31	56	07	08	28	50	46
51	47	46	64	99	68	10	72	36	21	94	04	99	13	45	42	83	60	91	91
99	55	96	83	31	62	53	52	41	70	69	77	71	28	30	74	81	97	81	42
41	46	88	51	49	49	55	41	79	94	14	92	43	96	50	95	29	40	05	56
94	55	93	75	59	49	67	85	31	19	70	31	20	56	82	66	98	63	40	99
41	61	57	03	60	64	11	45	86	60	90	85	06	46	18	80	62	05	17	90

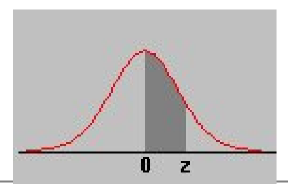
50	27	39	31	13	41	79	48	68	61	24	78	18	96	83	55	41	18	56	67
41	39	68	05	04	90	67	00	82	89	40	90	20	50	69	95	08	30	67	83
03	15	21	91	21	19	32	58	15	49	10	62	24	83	91	68	97	11	14	30
22	10	97	85	08	14	41	37	09	51	86	22	53	17	04	18	47	76	56	22
94	20	52	03	80	39	66	37	75	44	49	76	70	40	37	14	70	79	39	97
82	03	71	02	68	02	18	16	81	61	50	81	69	76	16	83	74	52	25	67
87	48	13	72	20	88	44	80	35	84	43	85	25	96	93	16	93	03	33	61
88	58	16	00	98	64	42	47	73	77	31	22	30	84	20	70	70	77	02	14
26	85	71	92	38	45	12	79	36	86	94	11	90	18	40	02	90	23	32	50
65	61	79	48	34	45	41	72	02	68	77	76	22	07	91	25	57	17	76	28
13	65	85	10	81	76	67	70	21	10	83	48	34	70	55	90	19	37	95	68
93	17	63	48	51	64	43	90	56	14	94	54	13	74	08	65	70	40	95	14
93	69	22	55	27	64	27	85	80	44	72	89	35	55	07	70	48	10	69	05

66	14	66	32	10	68	47	66	46	59	65	34	46	74	15	74	47	42	07	40
28	99	26	31	65	41	36	18	27	60	31	85	33	84	52	11	43	63	80	72
46	18	24	91	26	93	82	34	31	78	08	00	74	54	49	77	53	59	98	92
47	26	91	57	47	07	70	61	78	13	43	86	07	28	34	28	10	25	78	16
38	10	17	77	56	11	65	71	38	97	95	88	95	70	67	47	64	81	38	85
39	64	16	94	57	91	33	92	25	02	92	61	38	97	19	11	94	75	62	03
84	05	44	04	55	99	39	66	36	80	67	66	76	06	31	69	18	19	68	45
47	46	80	35	77	57	64	96	32	66	24	70	07	15	94	14	00	42	31	53
43	32	13	13	70	28	97	72	38	96	76	47	96	85	62	62	34	20	75	89
64	28	16	18	26	18	55	56	49	37	13	17	33	33	65	78	85	11	64	99
66	84	77	04	95	32	35	00	29	85	86	71	63	87	46	26	31	37	74	63
72	46	13	32	30	21	52	95	34	24	92	58	10	22	62	78	43	86	62	76
21	03	29	10	50	13	05	81	62	18	12	47	05	65	00	15	29	27	61	39
95	36	26	70	11	06	65	11	61	36	01	01	60	08	57	55	01	85	63	74

70	66	99	34	06	40	71	29	73	80	10	40	45	54	52	34	03	06	46	53
19	32	42	05	04	58	27	56	17	64	97	58	65	47	16	50	25	94	08	60
38	52	51	16	00	89	51	41	17	88	68	22	42	34	17	73	95	97	92	81
69	24	90	57	47	15	47	25	06	69	48	13	93	67	32	46	87	43	89	36
08	89	90	59	85	12	12	08	61	24	51	24	74	43	02	60	88	35	06	19
87	06	41	30	75	03	99	11	04	61	93	71	61	68	94	66	08	32	66	72
55	38	77	26	81	38	59	59	55	54	32	88	65	97	80	08	35	56	53	82
18	39	67	35	38	17	54	67	37	04	92	05	24	62	15	55	12	12	33	09
59	52	65	21	13	32	64	35	28	61	95	81	90	68	31	00	91	19	31	66
35	82	47	17	08	69	57	26	87	77	39	51	03	59	05	14	06	04	74	33

Area between 0 and *Z*

أخذت هذه البيانات في هذا الجدول من كتاب المبادئ الأولية في الإحصاء هو

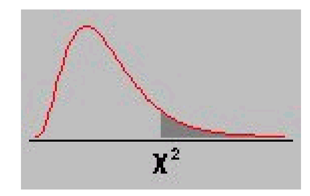


Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706

1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

أخذت هذه البيانات في هذا الجدول من كتاب " مبادئ الإحصاء لطلبة العلوم الإدارية والاقتصاد "تأليف هويل وجيسي

Right tail the areas for the χ^2 Distribution



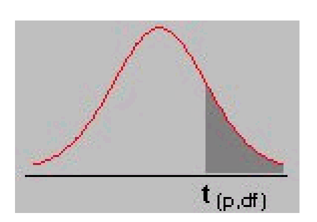
d.f	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.000	0.000	0.000	0.003	0.015	0.101	0.454	1.323	2.705	3.841	5.023	6.634	7.879
2	0.010	0.020	0.050	0.102	0.210	0.575	1.386	2.772	4.605	5.991	7.377	9.210	10.596
3	0.071	0.114	0.215	0.351	0.584	1.212	2.365	4.108	6.251	7.814	9.348	11.344	12.838
4	0.206	0.297	0.484	0.710	1.063	1.922	3.356	5.385	7.779	9.487	11.143	13.276	14.860
5	0.411	0.554	0.831	1.145	1.610	2.674	4.351	6.625	9.236	11.070	12.832	15.086	16.749
6	0.675	0.872	1.237	1.635	2.204	3.454	5.348	7.840	10.644	12.591	14.449	16.811	18.547
7	0.989	1.239	1.689	2.167	2.833	4.254	6.345	9.037	12.017	14.067	16.012	18.475	20.277
8	1.344	1.646	2.179	2.732	3.489	5.070	7.344	10.218	13.361	15.507	17.534	20.090	21.954
9	1.734	2.087	2.700	3.325	4.168	5.898	8.342	11.388	14.683	16.918	19.022	21.665	23.589
10	2.155	2.558	3.246	3.940	4.865	6.737	9.341	12.548	15.987	18.307	20.483	23.209	25.188

d.f	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
11	2.603	3.053	3.815	4.574	5.577	7.584	10.341	13.700	17.275	19.675	21.920	24.724	26.756
12	3.073	3.570	4.403	5.226	6.303	8.438	11.340	14.845	18.549	21.026	23.336	26.216	28.299
13	3.565	4.106	5.008	5.891	7.041	9.299	12.339	15.983	19.811	22.362	24.735	27.688	29.819
14	4.074	4.660	5.628	6.570	7.789	10.165	13.339	17.116	21.064	23.684	26.118	29.141	31.319
15	4.600	5.229	6.262	7.260	8.546	11.036	14.338	18.245	22.307	24.995	27.488	30.577	32.801
16	5.142	5.812	6.907	7.961	9.312	11.912	15.338	19.368	23.541	26.296	28.845	31.999	34.267
17	5.697	6.407	7.564	8.671	10.085	12.791	16.338	20.488	24.769	27.587	30.191	33.408	35.718
18	6.264	7.014	8.230	9.390	10.864	13.675	17.337	21.604	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.650	14.562	18.337	22.717	27.203	30.143	32.852	36.190	38.582
20	7.433	8.260	9.590	10.850	12.442	15.451	19.337	23.827	28.411	31.410	34.169	37.566	39.996
21	8.033	8.897	10.282	11.591	13.239	16.344	20.337	24.934	29.615	32.670	35.478	38.932	41.401
22	8.642	9.542	10.982	12.338	14.041	17.239	21.337	26.039	30.813	33.924	36.780	40.289	42.795
23	9.260	10.195	11.688	13.090	14.847	18.137	22.336	27.141	32.006	35.172	38.075	41.638	44.181
24	9.886	10.856	12.401	13.848	15.658	19.037	23.336	28.241	33.196	36.415	39.364	42.979	45.558

d.f	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
25	10.519	11.523	13.119	14.611	16.473	19.939	24.336	29.338	34.381	37.652	40.646	44.314	46.927
26	11.160	12.198	13.843	15.379	17.291	20.843	25.336	30.434	35.563	38.885	41.923	45.641	48.289
27	11.807	12.878	14.573	16.151	18.113	21.749	26.336	31.528	36.741	40.113	43.194	46.962	49.644
28	12.461	13.564	15.307	16.927	18.939	22.657	27.336	32.620	37.915	41.337	44.460	48.278	50.993
29	13.121	14.256	16.047	17.708	19.767	23.566	28.336	33.710	39.087	42.556	45.722	49.587	52.335
30	13.786	14.953	16.790	18.492	20.599	24.477	29.336	34.799	40.256	43.772	46.979	50.892	53.671

- أخذت هذه البيانات في هذا الجدول من كتاب " الإحصاء التطبيقي "تأليف دن وكلا

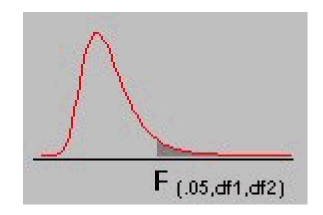
T table with right tail probabilities



d.f	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.32	1.00	3.07	6.31	12.70	31.82	63.65	636.61
2	0.28	0.81	1.88	2.91	4.30	6.96	9.92	31.59
3	0.27	0.76	1.63	2.35	3.18	4.54	5.84	12.92
4	0.27	0.74	1.53	2.13	2.77	3.74	4.60	8.61
5	0.26	0.72	1.47	2.01	2.57	3.36	4.03	6.86
6	0.26	0.71	1.43	1.94	2.44	3.14	3.70	5.95
7	0.26	0.71	1.41	1.89	2.36	2.99	3.49	5.40
8	0.26	0.70	1.39	1.85	2.30	2.89	3.35	5.04
9	0.26	0.70	1.38	1.83	2.26	2.82	3.24	4.78
10	0.26	0.69	1.37	1.81	2.22	2.76	3.16	4.58
11	0.25	0.69	1.36	1.79	2.20	2.71	3.10	4.43
12	0.25	0.69	1.35	1.78	2.17	2.68	3.05	4.31
13	0.25	0.69	1.35	1.77	2.16	2.65	3.01	4.22
14	0.25	0.69	1.34	1.76	2.14	2.62	2.97	4.14
15	0.25	0.69	1.34	1.75	2.13	2.60	2.94	4.07
16	0.25	0.69	1.33	1.74	2.11	2.58	2.92	4.01
17	0.25	0.68	1.33	1.73	2.10	2.56	2.89	3.96
18	0.25	0.68	1.33	1.73	2.10	2.55	2.87	3.92
19	0.25	0.68	1.32	1.72	2.09	2.53	2.86	3.88
20	0.25	0.68	1.32	1.72	2.08	2.52	2.84	3.84
21	0.25	0.68	1.32	1.72	2.07	2.51	2.83	3.81

d.f	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
22	0.25	0.68	1.32	1.71	2.07	2.50	2.81	3.79
23	0.25	0.68	1.31	1.71	2.06	2.49	2.80	3.76
24	0.25	0.68	1.31	1.71	2.06	2.49	2.79	3.74
25	0.25	0.68	1.31	1.70	2.05	2.48	2.78	3.72
26	0.25	0.68	1.31	1.70	2.05	2.47	2.77	3.70
27	0.25	0.68	1.31	1.70	2.05	2.47	2.77	3.68
28	0.25	0.68	1.31	1.70	2.04	2.46	2.76	3.67
29	0.25	0.68	1.31	1.69	2.04	2.46	2.75	3.65
30	0.25	0.68	1.31	1.69	2.04	2.45	2.75	3.64
Inf	0.25	0.67	1.28	1.64	1.95	2.32	2.57	3.29

- **أخذت هذه البيانات في هذا الجدول من كتاب " الإحصاء التطبيقي "تأليف دن وكلارك".**



F-Table for alpha=.05

DF2	1	2	3	4	5	6	7	8	9	10
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536

DF2	1	2	3	4	5	6	7	8	9	10
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	1.7306	1.6717

DF2	1	2	3	4	5	6	7	8	9	10
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105
INF	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307

DF2	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**INF**
1	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276

DF2	12	15	20	24	30	40	60	120	INF
9	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330

DF2	12	15	20	24	30	40	60	120	INF
25	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
INF	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

F-Table for alpha=.01

DF2/DF1	1	2	3	4	5	6	7	8	9	10
1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	5981.070	6022.473	6055.847
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805

DF2/DF1	1	2	3	4	5	6	7	8	9	10
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801

DF2/DF1	1	2	3	4	5	6	7	8	9	10
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472
INF	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321

DF2	12	15	20	24	30	40	60	120	INF
1	6106.321	6157.285	6208.730	6234.631	6260.649	6286.782	6313.030	6339.391	6365.864
2	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
6	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
11	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.602
12	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361

DF2	12	15	20	24	30	40	60	120	INF
13	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165	
14	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.868
16	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.653
18	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
21	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.305
23	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.256
24	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211
25	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.131
27	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
28	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064

DF2	12	15	20	24	30	40	60	120	INF
29	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
60	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
120	2.336	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.381
INF	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.325	1.000

جدول تحويل r إلى Z

(تحويل فشر لمعامل الارتباط)

R	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.0	.000	.010	.020	.030	.040	.050	.060	.070	.080	.090
.1	.100	.110	.121	.131	.141	.151	.161	.172	.182	.192
.2	.203	.213	.224	.234	.245	.255	.266	.277	.288	.299
.3	.310	.321	.332	.343	.354	.365	.377	.388	.400	.412
.4	.424	.436	.448	.460	.472	.485	.497	.510	.523	536
.5	.549	.563	.576	.590	.604	.618	.633	.648	.662	.678
.6	.693	.709	.725	.741	.758	.775	.793	.811	829	.848
.7	.867	.887	.908	.929	.950	.973	.996	1.020	1.045	1.071
.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422
r	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
.90	1.472	1.478	1.483	1.488	1.494	1.499	1.505	1.510	1.516	1.522
.91	1.528	1.533	1.539	1.545	1.551	1.557	1.564	1.570	1.576	1.583
.92	1.589	1.596	1.602	1.609	1.616	1.623	1.630	1.637	1.644	1.651
.93	1.658	1.666	1.673	1.681	1.689	1.697	1.705	1.713	1.721	1.730
.94	1.738	1.747	1.756	1.764	1.774	1.783	1.792	1.802	1.812	1.822
.95	1.832	1.842	1.853	1.863	1.874	1.886	1.897	1.909	1.921	1.933
.96	1.946	1.959	1.972	1.986	2.000	2.014	2.029	2.044	2.060	2.076
.97	2.092	2.109	2.127	2.146	2.165	2.185	2.205	2.227	2.249	2.273
.98	2.298	2.323	2.351	2.380	2.410	2.443	2.477	2.515	2.555	2.599
.99	2.646	2.700	2.759	2.826	2.903	2.994	3.106	3.250	3.453	3.800

❖ أخذت هذه البيانات في هذا الجدول من كتاب " الإحصاء التطبيقي "تأليف دن وكلارك"

جدول كولومجروف سيمنروف

لحساب القيم الحرجة لاختبار كولومجروف والتي تناظر

احتمال معين وحجم عينة معينة، وهو اختبار من طرف واحد

n	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.210	.220	.240	.270	.320
30	.190	.200	.220	.240	.290
35	.180	.190	.210	.230	.270
OVER 35	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

المصدر: أخذ الجدول من كتاب "مقدمة في الإحصاء التحليلي، تأليف" سمير كامل عاشور".

ملاحظة: إذا كانت $n > 30$ يعوض في آخر سطر بالجدول.